"Adaptive non-parametric instrumental regression in the presence of dependence"

Asin, Nicolas ; Johannes, Jan

Abstract

We consider the estimation of a structural function which models a nonparametric relationship between a response and an endogenous regressor given an instrument in presence of dependence in the data generating process. Assuming an independent and identically distributed (iid.) sample it has been shown in Johannes and Schwarz [2011] that a least squares estimator based on dimension reduction and thresholding can attain minimax-optimal rates of convergence up to a constant. As this estimation procedure requires an optimal choice of a dimension parameter with regard amongst others to certain characteristics of the unknown structural function we investigate its fully data-driven choice based on a combination of model selection and Lepski's method inspired by Goldenshluger and Lepski [2011]. For the resulting fully data-driven thresholded least squares estimator a non-asymptotic oracle risk bound is derived by considering either an iid. sample or by dismissing the independence assumptio...

Document type : *Document de travail (Working Paper)*

Référence bibliographique

Asin, Nicolas ; Johannes, Jan. *Adaptive non-parametric instrumental regression in the presence of dependence.* ISBA Discussion Paper ; 2016/15 (2016) 53 pages

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

DISCUSSION PAPER

2016/15

Adaptive non-parametric instrumental regression in the presence of dependence

Asin, N. and J. Johannes

Adaptive non-parametric instrumental regression in the presence of dependence

Nicolas Asin[1]

Université catholique de Louvain

Jan Johannes [2]

Ruprecht-Karls-Universität Heidelberg

Abstract

We consider the estimation of a structural function which models a non-parametric relationship between a response and an endogenous regressor given an instrument in presence of dependence in the data generating process. Assuming an independent and identically distributed (iid.) sample it has been shown in Johannes and Schwarz [2011] that a least squares estimator based on dimension reduction and thresholding can attain minimax-optimal rates of convergence up to a constant. As this estimation procedure requires an optimal choice of a dimension parameter with regard amongst others to certain characteristics of the unknown structural function we investigate its fully data-driven choice based on a combination of model selection and Lepski's method inspired by Goldenshluger and Lepski [2011]. For the resulting fully data-driven thresholded least squares estimator a non-asymptotic oracle risk bound is derived by considering either an iid. sample or by dismissing the independence assumption. In both cases the derived risk bounds coincide up to a constant assuming sufficiently weak dependence characterised by a fast decay of the mixing coefficients. Employing the risk bounds the minimax optimality up to constant of the estimator is established over a variety of classes of structural functions.

Keywords: Non-parametric regression, instrumental variable, dependence, mixing, minimax theory, adaptive.
JEL codes: C13, C14, C30, C36.

[1]ISBA, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, e-mail: nicolas.asin@uclouvain.be

[2]IAM, Ruprecht-Karls-Universität Heidelberg, Mathematikon, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany, e-mail: johannes@math.uni-heidelberg.de

1 Introduction

In non-parametric instrumental regression the relationship between a response Y and an endogenous explanatory variable Z is characterised by

$$Y = f(Z) + \varepsilon \quad \text{with} \quad \mathbb{E}(\varepsilon|Z) \neq 0 \tag{1.1a}$$

where the error term ε and Z are not stochastically mean-independent and f is called structural function. To account for the lack of mean-independence an additional exogenous random variable W, an instrument, is assumed, that is

$$\mathbb{E}(\varepsilon|W) = 0. \tag{1.1b}$$

In this paper we are interested in a fully data-driven estimation of the structural function f based on an identically distributed (id.) sample of (Y, Z, W) consisting either of independent or weakly dependent observations. Considering a thresholded least-squares estimator based on a dimension reduction with data-driven selection of the dimension parameter we show that the resulting fully data-driven estimator can attain optimal rates of convergence in a minimax sense.

Typical examples of models satisfying (1.1a–1.1b) are error-in-variable models, simultaneous equations or treatment models with endogenous selection. The natural generalisation (1.1a–1.1b) of a standard parametric model (e.g. Amemiya [1974]) to the non-parametric situation has been introduced by Florens [2003] and Newey and Powell [2003], while its identification has been studied e.g. in Carrasco et al. [2007], Darolles et al. [2011] and Florens et al. [2011]. Applications and extensions of this approach include non-parametric tests of exogeneity (Blundell and Horowitz [2007]), quantile regression models (Horowitz and Lee [2007]), semi-parametric modelling (Florens et al. [2012]), or quasi-Bayesian approaches (Florens and Simoni [2012]), to name but a few. There exists a vast literature on the non-parametric estimation of the structural function based on an iid. sample of (Y, Z, W). For example, Ai and Chen [2003], Blundell et al. [2007] or Newey and Powell [2003] consider sieve minimum distance estimators, Darolles et al. [2011], Florens et al. [2011] or Gagliardini and Scaillet [2012] study penalised least squares estimators, Dunker et al. [2014] propose an iteratively regularised Gauß–Newton methods, while iteratively regularised least squares estimators are analysed in Carrasco et al. [2007] and Johannes et al. [2013]. A least squares estimator based on dimension reduction and threshold techniques has been considered by Johannes and Schwarz [2011] and Breunig and Johannes [2015] which borrows ideas from the inverse problem community (c.f. Efromovich and Koltchinskii [2001] or Hoffmann and Reiß [2008]). Hall and Horowitz [2005], Chen and Reiß [2011] and Johannes and Schwarz [2011] prove lower bounds for the mean integrated squared error (MISE) and propose estimators which can attain optimal rates in a minimax sense. On the other hand lower bounds and minimax-optimal estimation of the value of a linear functional of the structural function has been shown in Breunig and Johannes [2015].

It is worth noting that all the proposed estimation procedures rely on the choice of at least one tuning parameter, which in turn, crucially influences the attainable accuracy of the constructed estimator. In general, this choice requires knowledge of characteristics of the structural function, such as the number of its derivatives, which are not known in practice. From an empirical point of view data-driven estimation procedures have been studied, for example, by Fève and Florens [2014], and Horowitz [2014]. Considering an iid. sample a fully data-driven estimation procedure for linear functionals of the structural function which can attain minimax-rates up to a logarithmic deterioration has been proposed by Breunig and Johannes [2015]. On the other hand side, based on an iid. sample data-driven estimators of the structural function which can attain lower bounds for the MISE are studied by Loubes and Marteau [2009] or Johannes and Schwarz [2011]. However, a straightforward application of their results is not obvious to us since they assume a partial knowledge of the associated conditional expectation of Z given W, that is, the eigenfunctions are known in advance, but the eigenvalues have to be estimated. In this paper we do not impose an a priori knowledge of the eigenbasis, and hence the estimators considered in Loubes and Marteau [2009] and Johannes and Schwarz [2011] are no more accessible to us. Instead, we consider a thresholded least squares estimator as presented in Johannes and Schwarz [2011].

Let us briefly sketch our fully data-driven estimation approach here. For the moment being, suppose that the structural function can be represented as $f = \sum_{j=1}^{m}[f]_j u_j$ using only m pre-specified basis functions $\{u_j\}_{j=1}^{m}$, and that only the coefficients $\{[f]_j\}_{j=1}^{m}$ with respect to these functions are unknown. In this situation, rewriting (1.1a–1.1b) as a multivariate linear conditional moment equation the estimation of the m coefficients of f is a classical textbook problem in econometrics (cf. Pagan and Ullah [1999]). A popular approach consists in replacing the conditional moment equation by an unconditional one, that is, $\mathbb{E}[Yv_l(W)] = \sum_{j=1}^{m}[f]_j\mathbb{E}[u_j(Z)v_l(W)]$, $l = 1, \dots, m$ given m functions $\{v_l\}_{l=1}^{m}$. Notice that once the functions $\{v_l\}_{l=1}^{m}$ are chosen, all the unknown quantities in the unconditional moment equations can be estimated by simply substituting empirical versions for the theoretical expectation. Moreover, a least squares solution of the estimated equation leads to a consistent and asymptotically normally distributed estimator of the coefficients vector of f under mild assumptions. The choice of the functions $\{v_l\}_{l=1}^{m}$ directly influences the asymptotic variance of the estimator and thus the question of optimal instruments minimising the asymptotic variance arises (cf. Newey

[1990]). However, in many situations an infinite number of functions $\{u_j\}_{j=1}^{\infty}$ and associated coefficients $\{[f]_j\}_{j=1}^{\infty}$ is needed to represent the structural function f, but we could still consider the finite dimensional least squares estimator described above for each dimension parameter $m \in \mathbb{N}$. In this situation the dimension m plays the role of a smoothing parameter and we may hope that the estimator of the structural function f is also consistent as m tends to infinity at a suitable rate. Unfortunately, this is not true in general. Let $f_m := \sum_{j=1}^{m}[f_m]_j u_j$ denote a least squares solution of the reduced unconditional moment equations, that is, the vector of coefficients $([f_m]_j)_{j=1}^{m}$ minimises the quantity $\sum_{l=1}^{m}\{\mathbb{E}[Yv_l(W)] - \sum_{j=1}^{m}\mathfrak{a}_j\mathbb{E}[u_j(Z)v_l(W)]\}^2$ over all $(\mathfrak{a}_j)_{j=1}^{m}$. Under an additional assumption (defined below) on the basis $\{v_j\}_{j\geqslant 1}$ it is shown in Johannes and Schwarz [2011] that f_m converges to the true structural function as m tends to infinity. Moreover, requiring a suitable chosen dimension parameter m a least squares estimator $\widehat{f}_m$ of f based on a dimension reduction together with an additional thresholding can attain minimax-optimal rates of convergence in terms of the MISE. In this paper we make use of a method to select the dimension parameter in a fully data-driven way, that is, neither depending on the structural function nor on the underlying joint distribution of Z and W. Inspired by the work of Goldenshluger and Lepski [2011] the procedure combines a model selection approach (cf. Barron et al. [1999] and its detailed discussion in Massart [2007]) and Lepski's method (cf. Lepski [1990]).

The main contribution of this paper is the derivation of a non-asymptotic oracle bound of the MISE for the resulting fully data-driven thresholded least squares estimator by considering either an iid. sample or by dismissing the independence assumption. Employing these bounds the minimax optimality up to constant of the estimator is established in terms of the MISE over a variety of classes of structural functions and conditional expectations. The estimator which depends only on the data adapts thus automatically to the unknown characteristics of the structural function.

The paper is organised as follows: in Section 2 we introduce our basic model assumptions and notations, introduce the thresholded least squares estimator $\widehat{f}_m$ as proposed in Johannes and Schwarz [2011] and present the data-driven method to select the tuning parameter $\widehat{m}$. We prove in Section 3 an oracle upper bound of the MISE for the resulting fully data-driven estimator $\widehat{f}_{\widehat{m}}$ assuming first that the id. sample of (Y, Z, W) consists of independent observations and second that the sample is drawn from a strictly stationary process. We briefly review elementary dependence notions and present standard coupling arguments. The risk bounds are non-asymptotic and depend as usual on the structural function and the conditional expectation. Employing these risk bounds we show in Section 4 that within the general framework as presented in Johannes and Schwarz [2011] the fully data-driven estimator $\widehat{f}_{\widehat{m}}$ can attain up to a constant the lower bound of the maximal MISE over a variety of classes of structural functions and conditional expectations. In particular we provide sufficient conditions on the dependence structure such that the fully data-driven estimator based on the dependent observations can still attain the minimax-rates for independent data.

2 Assumptions and methodology

Basic model assumptions For ease of presentation we consider a scalar regressor Z and a scalar instrument W. However, all the results below can be extended to the multivariate case in a straightforward way. It is convenient to rewrite the model (1.1a–1.1b) in terms of an operator between Hilbert spaces. Let us first introduce the Hilbert spaces $L^2_Z := \{f : \mathbb{R} \to \mathbb{R} \,|\, \|f\|^2_Z := \mathbb{E}[f^2(Z)] < \infty\}$ and $L^2_W := \{g : \mathbb{R} \to \mathbb{R} \,|\, \|g\|^2_W := \mathbb{E}[g^2(W)] < \infty\}$ endowed with the usual inner products $\langle\cdot,\cdot\rangle_Z$ and $\langle\cdot,\cdot\rangle_W$, respectively. For the sake of simplicity and ease of understanding, we follow and refer the reader to Hall and Horowitz [2005] for a discussion of the assumption that Z and W are marginally uniformly distributed on the interval $[0, 1]$. Obviously, in this situation both Hilbert spaces L^2_Z and L^2_W are isomorphic to $L^2 := L^2[0, 1]$ endowed with the usual norm $\|\cdot\|_{L^2}$ and inner product $\langle\cdot,\cdot\rangle_{L^2}$. The conditional expectation of Z given W, however, defines a linear operator $Tf := \mathbb{E}[f(Z)|W]$, $f \in L^2_Z$ mapping L^2_Z into L^2_W. Taking the conditional expectation with respect to the instrument W on both sides in (1.1a) we obtain from (1.1b) that:

$$g := \mathbb{E}(Y|W) = \mathbb{E}(f(Z)|W) =: Tf \tag{2.1}$$

where the function g belongs to L^2_W. Estimation of the structural function f is thus linked to the inversion of T and it is therefore called an inverse problem. Here und subsequently, we suppose implicitly that the operator T is compact, which is the case under fairly mild assumptions (c.f. Carrasco et al. [2006]). Consequently, unlike in a multivariate linear instrumental regression model, a continuous generalised inverse of T does not exist as long as the range of the operator T is an infinite dimensional subspace of L^2_W. This corresponds to the set-up of ill-posed inverse problems with the additional difficulty that T is unknown and has to be estimated. In what follows, it is always assumed that there exists a unique solution $f \in L^2_Z$ of equation (2.1), in other words, that g belongs to the range of T, and that T is injective. For a detailed discussion in the context of inverse problems see Chapter 2.1 in Engl et al. [2000], while in the special case of a non-parametric instrumental regression we refer to Carrasco et al. [2006]. Considering $\mu(Z, W) := \mathbb{E}[\varepsilon|Z, W]$ we decompose throughout the paper the error term $\varepsilon = \xi + \mu(Z, W)$ where ξ is centred due to the mean independence of ε given the instrument W as supposed in (1.1b). Moreover, we assume that ξ and (Z,W) are independent of each other. Denoting by $\|h\|_\infty$ and $\|h\|_{Z,W} := (\mathbb{E}h^2(Z,W))^{1/2}$, respectively, the usual uniform norm and L^2-norm of a real valued function h the next assumption completes and formalises our conditions on the regressor Z, the instrument W and the random variable ξ.

Assumption A.1. *The joint distribution of (Z,W) admits a bounded density $p_{Z,W}$, i.e., $\|p_{Z,W}\|_\infty < \infty$, while both Z and W are marginally uniformly distributed on the interval $[0,1]$. The conditional mean function $\mu(Z,W) := \mathbb{E}[\varepsilon|Z,W]$ is uniformly bound, that is, $\|\mu\|_\infty < \infty$ and, thus $\|\mu\|_{Z,W} < \infty$. The random variables $\{\xi_i := \varepsilon_i - \mu(Z_i,W_i)\}_{i=1}^n$ form an iid. n-sample of $\xi := \varepsilon - \mu(Z,W)$ satisfying $\mathbb{E}\xi^{12} < \infty$ and $\sigma_\xi^2 := \mathbb{E}\xi^2 > 0$, which is independent of $\{(Z_i,W_i)\}_{i=1}^n$.*

Matrix and operator notations We base our estimation procedure on the expansion of the structural function f and the conditional expectation operator T in an orthonormal basis of L_Z^2 and L_W^2, respectively. The selection of an adequate basis in non-parametric instrumental regression, and inverse problems in particular, is discussed in various publications, (c.f. Efromovich and Koltchinskii [2001] or Breunig and Johannes [2015], and references within). We may emphasise that, the basis in L_Z^2 is determined by the presumed information on the structural function and is not necessarily an eigenbasis for the unknown operator. However, the statistical choice of a basis from a family of bases (c.f. Birgé and Massart [1997]) is complicated, and its discussion is far beyond the scope of this paper. Therefore, we assume here and subsequently that $\{u_j\}_{j=1}^\infty$ and $\{v_j\}_{j=1}^\infty$ denotes an adequate orthonormal basis of L_Z^2 and L_W^2, respectively, which do not in general correspond to the eigenfunctions of the operator T defined in (2.1). The next assumption summarises our minimal conditions on those basis.

Assumption A.2. *There exists a finite constant $\tau_\infty^2 \geqslant 1$ such that the basis $\{u_j\}_{j=1}^\infty$ and $\{v_j\}_{j=1}^\infty$ satisfy $\|\sum_{j=1}^m u_j^2\|_\infty \leqslant m\tau_\infty^2$ and $\|\sum_{j=1}^m v_j^2\|_\infty \leqslant m\tau_\infty^2$, for any $m \in \mathbb{N}$.*

According to Lemma 6 of Birgé and Massart [1997] Assumption A.2 is exactly equivalent to following property: there exists a positive constant τ_∞ such that for any h belongs to the subspace $\mathbb{D}_m$, spanned by the first m basis functions, holds $\|h\|_\infty \leqslant \tau_\infty\sqrt{m}\|h\|_Z$. Typical example are bounded basis, such as the trigonometric basis, or basis satisfying the assertion, that there exists a positive constant C_∞ such that for any $(c_1,\dots,c_m) \in \mathbb{R}^m$, $\|\sum_{j=1}^m c_j u_j\|_\infty \leqslant C_\infty\sqrt{m}|c|_\infty$ where $|c|_\infty = \max_{1\leqslant j\leqslant m} c_j$. Birgé and Massart [1997] have shown that the last property is satisfied for piece-wise polynomials, splines and wavelets.

Given the orthonormal basis $\{u_j\}_{j=1}^\infty$ and $\{v_j\}_{j=1}^\infty$ of L_Z^2 and L_W^2, respectively, we consider for all $f \in L_Z^2$ and $g \in L_W^2$ the development $f = \sum_{j=1}^\infty [f]_j u_j$ and $g = \sum_{j=1}^\infty [g]_j v_j$

where with a slight abuse of notation the sequences $([f]_j)_{j\geqslant 1}$ and $([g]_j)_{j\geqslant 1}$ with generic elements $[f]_j := \langle f, u_j\rangle_Z$ and $[g]_j := \langle g, v_j\rangle_W$ are square-summable, that is, $\|f\|_Z^2 = \sum_{j=1}^\infty [f]_j^2 < \infty$ and $\|g\|_W^2 = \sum_{j=1}^\infty [g]_j^2 < \infty$. We will refer to any sequence as a whole by omitting its index as for example in «the sequence $[f]$». Furthermore, for $m \geqslant 1$ let $[f]_{\underline{m}} := ([f]_j, \dots, [f]_m)^t$ (resp. $[g]_{\underline{m}}$) where x^t is the transpose of x. Let us further denote by $\mathcal{U}_m$ and $\mathcal{V}_m$ the subspace of L_Z^2 and L_W^2 spanned by the basis functions $\{u_j\}_{j=1}^m$ and $\{v_j\}_{j=1}^m$, respectively. Obviously, the norm of $f \in \mathcal{U}_m$ equals the euclidean norm of its coefficient vector $[f]_{\underline{m}}$, that is, $\|f\|_Z = ([f]_{\underline{m}}^t [f]_{\underline{m}})^{1/2} =: \|[f]_{\underline{m}}\|$. Clearly, if (Y, Z, W) obeys the model equations (1.1a–1.1b) then introducing the infinite dimensional random vector $[v(W)]$ with generic elements $[v(W)]_j = v_j(W)$ the identity $[g]_{\underline{m}} := \mathbb{E}(Y[v(W)]_{\underline{m}})$ holds true. Consider in addition the infinite dimensional random vector $[u(Z)]$ with generic elements $[u(Z)]_j = u_j(Z)$. We define the $m \times m$ dimensional matrix $[T]_{\underline{m}} := \mathbb{E}([v(W)]_{\underline{m}}[u(Z)]_{\underline{m}}^t)$ with generic elements $\langle v_l, Tu_j\rangle_W$ which is throughout the paper assumed to be non singular for all $m \geqslant 1$ (or, at least for sufficiently large m), so that $[T]_{\underline{m}}^{-1}$ always exists with finite spectral norm $\|[T]_{\underline{m}}^{-1}\|_s := \sup_{\|v\|\leqslant 1}\|[T]_{\underline{m}}^{-1} v\| < \infty$. Note that it is a non-trivial problem to determine under what precise conditions such an assumption holds (see e.g. Efromovich and Koltchinskii [2001] and references therein). We consider the approximation $f_m \in \mathcal{U}_m$ of f given by $[f_m]_{\underline{m}} = [T]_{\underline{m}}^{-1}[g]_{\underline{m}}$ and $[f_m]_j = 0$ for all $j > m$. Although, it does generally not correspond to the orthogonal projection of f onto the subspace $\mathcal{U}_m$ and the approximation error $\mathfrak{b}_m^2(f) := \sup_{k\geqslant m}\|f_m - f\|_Z^2$ does generally not converge to zero as $m \to \infty$. Here and subsequently, however, we restrict ourselves to cases of structural functions and conditional expectation operators which ensure the convergence. Obviously, this is a minimal regularity condition for us since we aim to estimate the approximation f_m.

Thresholded least squares estimator In this paper, we follow Johannes and Schwarz [2011] and consider a least squares solution of a reduced set of unconditional moment equations which takes its inspiration from the linear Galerkin approach used in the inverse problem community (c.f. Efromovich and Koltchinskii [2001] or Hoffmann and Reiß [2008]). To be precise, let $\{(Y_i, Z_i, W_i)\}_{i=1}^n$ be an identically distributed sample of (Y, Z, W) obeying (1.1a–1.1b). Since $[T]_{\underline{m}} = \mathbb{E}[v(W)]_{\underline{m}}[u(Z)]_{\underline{m}}^t$ and $[g]_{\underline{m}} = \mathbb{E}Y[v(W)]_{\underline{m}}$ are written as expectations we can construct estimators using their empirical counterparts, that is, $\widehat{[T]}_{\underline{m}} := n^{-1}\sum_{i=1}^n [v(W_i)]_{\underline{m}}[u(Z_i)]_{\underline{m}}^t$ and $\widehat{[g]}_{\underline{m}} := n^{-1}\sum_{i=1}^n Y_i[v(W_i)]_{\underline{m}}$. Let $\mathbb{1}_{\{\|\widehat{[T]}_{\underline{m}}^{-1}\|_s^2 \leqslant n\}}$ denote the indicator function which takes the value one if $\widehat{[T]}_{\underline{m}}$ is non singular with squared spectral norm $\|\widehat{[T]}_{\underline{m}}^{-1}\|_s^2$ bounded by n. The estimator $\widehat{f}_m \in \mathcal{U}_m$ of the structural function f is then defined by

$$[\widehat{f}_m]_{\underline{m}} := \widehat{[T]}_{\underline{m}}^{-1}\widehat{[g]}_{\underline{m}}\mathbb{1}_{\{\|\widehat{[T]}_{\underline{m}}^{-1}\|_s \leqslant n\}} \tag{2.2}$$

where the dimension parameter $m = m(n)$ has to tend to infinity as the sample size n increases.

Data-driven dimension selection Our selection method combines model selection (c.f. Barron et al. [1999] and its discussion in Massart [2007]) and Lepskij's method (c.f. Lepski [1990]) borrowing ideas from Goldenshluger and Lepski [2011]. We select the dimension parameter as minimiser of a penalised contrast function which we formalise next. Given a positive sequence $\mathfrak{a} := (\mathfrak{a}_m)_{m\geqslant 1}$ denote

$$\Delta_m(\mathfrak{a}) := \max_{1\leqslant k\leqslant m} \mathfrak{a}_k, \quad \Lambda_m(\mathfrak{a}) := \max_{1\leqslant k\leqslant m} \frac{\log(\mathfrak{a}_k \vee (k+2))}{\log(k+2)} \quad \text{and} \\ \delta_m(\mathfrak{a}) := m\, \Delta_m(\mathfrak{a})\, \Lambda_m(\mathfrak{a}). \quad (2.3)$$

Thereby, we define $\widehat{\delta}_m := \delta_m(\mathfrak{a})$ with $\mathfrak{a} = (\|[\widehat{T}]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}$. For $n \geqslant 1$, a positive sequence $\mathfrak{a} := (\mathfrak{a}_m)_{m\geqslant 1}$ and $\alpha_n := n^{1-1/\log(2+\log n)}(1+\log n)^{-1}$ denote

$$\mathrm{M}_n(\mathfrak{a}) := \min\Big\{2 \leqslant m \leqslant \lfloor n^{1/4} \rfloor : m^2\, \mathfrak{a}_m > \alpha_n\Big\} - 1 \quad (2.4)$$

where we set $\mathrm{M}_n(\mathfrak{a}) := \lfloor n^{1/4} \rfloor$ if the minimum is taken over an empty set and $\lfloor x \rfloor$ denotes as usual the integer part of x. Thereby, the dimension parameter is selected among a collection of admissible values $\{1, \dots, \widehat{\mathrm{M}}\}$ with random integer $\widehat{\mathrm{M}} = \mathrm{M}_n(\mathfrak{a})$ and $\mathfrak{a} = (\|[\widehat{T}]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}$. Taking its inspiration from Comte and Johannes [2012] the stochastic sequence of penalties $(\widehat{\mathrm{pen}}_m)_{1\leqslant m\leqslant \widehat{\mathrm{M}}}$ is defined by

$$\widehat{\mathrm{pen}}_m := 11\, \kappa\, \widehat{\sigma}_m^2\, \widehat{\delta}_m\, n^{-1} \quad \text{with} \quad \widehat{\sigma}_m^2 := 2\Big(\sum_{i=1}^n Y_i^2 + \max_{1\leqslant k\leqslant m} \|\widehat{f}_k\|_Z^2\Big) \quad (2.5)$$

where κ is a positive constant to be chosen below. The random integer $\widehat{\mathrm{M}}$ and the stochastic penalties $(\widehat{\mathrm{pen}}_m)_{1\leqslant m\leqslant \widehat{\mathrm{M}}}$ are used to define the sequence of contrasts $(\widehat{\Upsilon}_m)_{1\leqslant m\leqslant \widehat{\mathrm{M}}}$ by

$$\widehat{\Upsilon}_m := \max_{m\leqslant k\leqslant \widehat{\mathrm{M}}} \Big\{\|\widehat{f}_k - \widehat{f}_m\|_Z^2 - \widehat{\mathrm{pen}}_k\Big\}. \quad (2.6)$$

Setting $\arg\min_{m\in A}\{\mathfrak{a}_m\} := \min\{m : \mathfrak{a}_m \leqslant \mathfrak{a}_{m'},\ \forall m' \in A\}$ for a sequence $(\mathfrak{a}_m)_{m\geqslant 1}$ with minimal value in $A \subset \mathbb{N}$ we select the dimension parameter

$$\widehat{m} := \mathop{\arg\min}_{1\leqslant m\leqslant \widehat{\mathrm{M}}} \Big\{\widehat{\Upsilon}_m + \widehat{\mathrm{pen}}_m\Big\}. \quad (2.7)$$

The estimator of f is now given by $\widehat{f}_{\widehat{m}}$ and below we derive an upper bound for its risk $\mathbb{E}\|\widehat{f}_{\widehat{m}} - f\|_Z^2$. By construction the choice of the dimension parameter and hence the estimator $\widehat{f}_{\widehat{m}}$ do rely neither on the structural function and the conditional expectation operator nor on their regularity assumptions which we formalise in Section 4.

3 Non asymptotic oracle risk bound

3.1 Independent observations

In this section we derive an upper bound for the MISE of the thresholded least squares estimator $\widehat{f}_{\widehat{m}}$ with data-driven choice $\widehat{m}$ of the dimension parameter. We first suppose that the identically distributed n-sample $\{(Y_i, Z_i, W_i)\}_{i=1}^n$ consists of independent random variables. In a second step we dismiss below the independence assumption by imposing that $\{(Z_i, W_i)\}_{i=1}^n$ are weakly dependent. The next assumption summarises our conditions on the operator, the solution and its approximation.

Assumption A.3. *(a) The matrix $[T]_{\underline{m}}$ is non singular for all $m \geqslant 1$ such that $[T]_{\underline{m}}^{-1}$ always exists.*

(b) The function μ as in Assumption A.1, the structural function f and its approximation $f_m \in \mathcal{U}_m$ given by $[f_m]_{\underline{m}} = [T]_{\underline{m}}^{-1}[g]_{\underline{m}}$ satisfy $\|\mu\|_{Z,W}^2 \vee \|f\|_Z^2 \vee \sup_{m\geqslant 1}\|f_m\|_Z^2 \leqslant \Gamma_2^f < \infty$ and $\|\mu\|_\infty + \|f\|_\infty + \sup_{m\geqslant 1}\|f - f_m\|_\infty \leqslant \Gamma_\infty^f < \infty$.

The formulation of the upper risk bound relies on theoretical counterparts to the random quantities $\widehat{M}$ and $\widehat{\mathrm{pen}}_m$ which amongst other we define now referring only to the structural function f and the operator T. Keep in mind the notation given in (2.3) and (2.4). For $m, n \geqslant 1$ and $\mathfrak{a} := (\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}$ define $\Delta_m^T := \Delta_m(\mathfrak{a})$, $\Lambda_m^T := \Lambda_m(\mathfrak{a})$ and $\delta_m^T := m\Delta_m^T\Lambda_m^T$, set $\mathrm{M_n^{T-}} := \mathrm{M}_n(4\mathfrak{a})$ and $\mathrm{M_n^{T+}} := \mathrm{M}_n(\mathfrak{a}/4)$ where $\mathrm{M_n^{T-}} \leqslant \mathrm{M_n^{T+}}$ by construction. We require in addition that the sequence $(\mathrm{M_n^{T+}})_{n\geqslant 1}$ satisfies $\log(n)(\mathrm{M_n^{T+}}+1)^2\Delta_{\mathrm{M_n^{T+}}+1}^T = o(n)$ as $n \to \infty$. In Section 4.2 below we provide an Illustration considering different configurations for the decay of the sequence $(\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}$ where this condition is automatically satisfied.

Theorem 3.1. *Assume an i.i.d. n-sample of (Y, Z, W) obeying (1.1a–1.1b). Let Assumption A.1, A.2 and A.3 be satisfied. Set $\kappa = 144$ in the definition (2.5) of the penalty $\widehat{\mathrm{pen}}_m$. If $\log(n)(\mathrm{M_n^{T+}}+1)^2\Delta_{\mathrm{M_n^{T+}}+1}^T = o(n)$ as $n \to \infty$, then there exists a constant Σ^f given as in (C.3) in the Appendix C, which depends amongst others on τ_∞, Γ_∞^f and σ_ξ, and a numerical constant C such that for all $n \geqslant 1$*

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2\Big) \leqslant C\,\tau_\infty^2(1 + \sigma_\xi^2 + \Gamma_2^f)\{\min_{1\leqslant m\leqslant \mathrm{M_n^{T-}}}\{[\mathfrak{b}_m^2(f) \vee \delta_m^T n^{-1}]\} + n^{-1}\,\Sigma^f\}.$$

Let us briefly comment on the last result. We shall emphasise that the derived upper bound holds for all $n \geqslant 1$ true and thus is non-asymptotic. The bound consists of two terms, a remainder term $n^{-1}\,\Sigma^f$ which is negligible with respect to the first rhs term $\min_{1\leqslant m\leqslant \mathrm{M_n^{T-}}}\{[\mathfrak{b}_m^2(f) \vee \delta_m^T n^{-1}]\}$. The dependence of the factor Σ^f in the remainder term on the unknown structural function f (and the conditional expectation operator T) is explicitly given in its definition (C.3). This dependence is rather complicated but allows us still to derive in the next section an uniform bound of Σ^f over certain classes of structural functions and conditional expectation operators. On the other hand side, identifying for $1 \leqslant m \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}$, $\mathfrak{b}_m^2(f)$ as upper bound of the squared-bias and $\delta_m^T n^{-1}$ as upper bound of the variance of the thresholded least squares estimator $\widehat{f}_m$ the dominating term $\min_{1 \leqslant m \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}}\{[\mathfrak{b}_m^2(f) \vee \delta_m^T n^{-1}]\}$ mimics a squared-bias-variance trade-off. Let us further introduce

$$m_n^\diamond := \underset{1 \leqslant m \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}}{\arg\min}\{[\mathfrak{b}_m^2(f) \vee \delta_m^T n^{-1}]\} \quad \text{and} \quad \mathcal{R}_n^\diamond := [\mathfrak{b}_{m_n^\diamond}^2(f) \vee \delta_{m_n^\diamond}^T n^{-1}]. \tag{3.1}$$

Obviously, the estimator $\widehat{f}_{m_n^\diamond}$ minimises within the family $\{\widehat{f}_1, \dots, \widehat{f}_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}}\}$ of estimators the upper bound for the risk. The dimension parameter $m_n^\diamond$ and, hence the estimator $\widehat{f}_{m_n^\diamond}$ depend, however, on the unknown structural function and conditional expectation operator. The estimator $\widehat{f}_{m_n^\diamond}$ is therefore not feasible, and called an oracle. We shall emphasise that due to Theorem 3.1 the risk of the data-driven estimator $\widehat{f}_{\widehat{m}}$ is bounded up to a constant by the risk $\mathcal{R}_n^\diamond$ of the oracle within the family $\{\widehat{f}_1, \dots, \widehat{f}_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}}\}$. Moreover, we will show in Section 4 below that $\mathcal{R}_n^\diamond$ is the minimax-optimal rate for a wide range of classes of structural functions and conditional expectation operators which in turn establishes minimax optimality of the data-driven estimator.

3.2 Dependent observations

In this section we dismiss the independence assumption and assume weakly dependent observations. More precisely, $(Z_1, W_1), \dots, (Z_n, W_n)$ are drawn from a strictly stationary process $\{(Z_i, W_i)\}_{i \in \mathbb{Z}}$. Keep in mind that a process is called strictly stationary if its finite dimensional distributions do not change when shifted in time. We suppose that the observations $\{(Y_i, Z_i, W_i)\}_{i=1}^n$ still form an identically distributed sample from (Y, Z, W) obeying the model (1.1a–1.1b). Our aim is the non-parametric estimation of the structural function f under some mixing conditions on the dependence of the process $\{(Z_i, W_i)\}_{i \in \mathbb{Z}}$. Let us begin with a brief review of a classical measure of dependence, leading to the notion of a stationary absolutely regular process.

Let $(\Omega, \mathscr{A}, P)$ be a probability space. Given two sub-σ-fields $\mathscr{U}$ and $\mathscr{V}$ of $\mathscr{A}$ we introduce next the definition and properties of the absolutely regular mixing (or β-mixing) coefficient $\beta(\mathscr{U}, \mathscr{V})$. The coefficient was introduced by Kolmogorov and Rozanov [1960] and is defined by

$$\beta(\mathscr{U}, \mathscr{V}) := \tfrac{1}{2} \sup \left\{ \sum_i \sum_j |P(U_i)P(V_i) - P(U_i \cap V_i)| \right\}$$

where the supremum is taken over all finite partitions $(U_i)_{i \in I}$ and $(V_j)_{j \in J}$, which are respectively $\mathscr{U}$ and $\mathscr{V}$ measurable. Obviously, $\beta(\mathscr{U}, \mathscr{V}) \leqslant 1$. As usual, if U and U' are two random variables defined on $(\Omega, \mathscr{A}, P)$, we denote by $\beta(U, U')$ the mixing coefficient $\beta(\sigma(U), \sigma(U'))$, where $\sigma(U)$ and $\sigma(U')$ are, respectively, the σ-fields generated by U and U'.

We assume in the sequel that there exists a sequence of independent random variables with uniform distribution on $[0,1]$ independent of the strictly stationary process $\{(Z_i, W_i)\}_{i\in\mathbb{Z}}$. Employing Lemma 5.1 in Viennet [1997] we construct by induction a process $\{(Z_i^\perp, W_i^\perp)\}_{i\geqslant 1}$ satisfying the following properties. Given an integer q we introduce disjoint even and odd blocks of indices, i.e., for any $l \geqslant 1$, $\mathcal{I}_l^e := \{2(l-1)q+1, \dots, (2l-1)q\}$ and $\mathcal{I}_l^o := \{(2l-1)q+1, \dots, 2lq\}$, respectively, of size q. Let us further partition into blocks the random processes $\{(Z_i, W_i)\}_{i\geqslant 1} = \{(E_l, O_l)\}_{l\geqslant 1}$ and $\{(Z_i^\perp, W_i^\perp)\}_{i\geqslant 1} = \{(E_l^\perp, O_l^\perp)\}_{l\geqslant 1}$ where

$$E_l = (Z_i, W_i)_{i\in\mathcal{I}_l^e}, \quad E_l^\perp = (Z_i^\perp, W_i^\perp)_{i\in\mathcal{I}_l^e}, \quad O_l = (Z_i, W_i)_{i\in\mathcal{I}_l^o}, \quad O_l^\perp = (Z_i^\perp, W_i^\perp)_{i\in\mathcal{I}_l^o}.$$

If we set further $\mathscr{F}_l^- := \sigma((Z_j, W_j), j \leqslant l)$ and $\mathscr{F}_l^+ := \sigma((Z_j, W_j), j \geqslant l)$, then the sequence $(\beta_k)_{k\geqslant 0}$ of β-mixing coefficient defined by $\beta_0 := 1$ and $\beta_k := \beta(\mathscr{F}_0^-, \mathscr{F}_k^+)$, $k \geqslant 1$, is monotonically non-increasing and satisfies trivially $\beta_k \geqslant \beta((Z_0, W_0), (Z_k, W_k))$ for any $k \geqslant 1$. Based on the construction presented in Viennet [1997], the sequence $(Z_i^\perp, W_i^\perp)_{i\geqslant 1}$ can be chosen such that for any integer $l \geqslant 1$:

(P1) $E_l^\perp$, E_l, $O_l^\perp$ and O_l are identically distributed,

(P2) $P(E_l \neq E_l^\perp) \leqslant \beta_{q+1}$, and $P(O_l \neq O_l^\perp) \leqslant \beta_{q+1}$.

(P3) The variables $(E_1^\perp, \dots, E_l^\perp)$ are iid. and so $(O_1^\perp, \dots, O_l^\perp)$.

We shall emphasise that the random vectors $E_1^\perp, \dots, E_l^\perp$ are iid. but the components within each vector are generally not independent. The next result requires the following assumption which has been used, for example, in Bosq [1998].

Assumption A.4. *For any integer k the joint distribution P_{Z_0,W_0,Z_k,W_k} of (Z_0, W_0) and (Z_k, W_k) admits a density p_{Z_0,W_0,Z_k,W_k} which is square integrable and satisfies* $\Gamma_{ZW} := \sup_{k\geqslant 1}\|p_{(Z_0,W_0),(Z_k,W_k)} - p_{Z,W} \otimes p_{Z,W}\|_{Z,W\times Z,W} < \infty$.

Theorem 3.2. *Assume a sample $\{(Y_i, Z_i, W_i)\}_{i=1}^n$ obeying (1.1a–1.1b) where $\{(Z_i, W_i)\}_{i=1}^n$ is drawn from a stationary absolutely regular process with mixing coefficients $(\beta_k)_{k\geqslant 0}$ satisfying $\mathfrak{B} := \sum_{k=0}^\infty (k+1)^2\beta_k < \infty$ and given $k \geqslant 1$ set $\mathfrak{B}_k := \sum_{j=k}^\infty \beta_k \leqslant \mathfrak{B}$. Let the Assumptions A.1–A.4 be satisfied. Considering the oracle dimension $m_n^\diamond$ as in (3.1) let $k_n := \lfloor (\Gamma_\infty^f/\sigma_\xi)^2 \Gamma_{ZW} m_n^\diamond \rfloor$ and $\kappa_n^f \in [6 + 8(\Gamma_\infty^f/\sigma_\xi)^2 \mathfrak{B}_{k_n}, 8(1 + (\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B})]$. Set $\kappa = 288\kappa_n^f$ in the definition (2.5) of the penalty $\widehat{\mathrm{pen}}_m$. If $\log(n)(\mathrm{M}_n^{\mathrm{T}+}+1)^2 \Delta^T_{\mathrm{M}_n^{\mathrm{T}+}+1} = o(n)$ as $n \to \infty$, then there exists a constant Σ^f given as in (D.3) in the Appendix D, which* *depends amongst others on τ_∞, Γ^f_∞, σ_ξ and $\mathfrak{B}$, and a numerical constant C such that for all $1 \leqslant q \leqslant n$*

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2\Big) \leqslant C\,\Big\{[\mathfrak{b}^2_{m_n^\diamond}(f) \vee n^{-1}\delta^T_{m_n^\diamond}] + n^{-1}[\Sigma^f \vee n^3\exp(-n^{1/6}q^{-1}/100) \vee n^4 q^{-1}\beta_{q+1}]\Big\} \\ \times\, \tau_\infty^2(1+\sigma_\xi^2+\Gamma_2^f)(1+(\Gamma^f_\infty/\sigma_\xi)^2\mathfrak{B}).$$

We shall emphasise that the last assertion provides again a non-asymptotic risk bound for the estimator $\widehat{f}_{\widehat{m}}$ with dimension $\widehat{m}$ as in (2.7). Note that, the quantity κ_n^f used to construct the penalty $\widehat{\mathrm{pen}}_m$ in the last theorem still depends on the mixing coefficients $(\beta_k)_{k\geqslant 0}$ which are generally unknown. However, the condition $\mathfrak{B} = \sum_{k=0}^\infty (k+1)^2\beta_k < \infty$ implies $\sum_{k=k_n}^\infty \beta_k \leqslant (k_n+1)^{-2}\mathfrak{B}$ and hence, $6 + 8(\Gamma^f_\infty/\sigma_\xi)^2\mathfrak{B}_{k_n} \leqslant 7$ whenever $k_n = \lfloor(\Gamma^f_\infty/\sigma_\xi)^2\Gamma_{ZW} m_n^\diamond\rfloor \geqslant \sqrt{8(\Gamma^f_\infty/\sigma_\xi)^2\mathfrak{B}_{k_n}}$. Thereby, if $m_n^\diamond \to \infty$ as $n \to \infty$, then there exists an integer n_o such that $\kappa_n^f = 7 \in [6 + 8(\Gamma^f_\infty/\sigma_\xi)^2\mathfrak{B}_{k_n}, 8(1+(\Gamma^f_\infty/\sigma_\xi)^2\mathfrak{B})]$ for all $n \geqslant n_o$. The next assertion is thus an immediate consequence of Theorem 3.2, and hence its proof is omitted.

Corollary 3.3. *Let the assumptions of Theorem 3.2 be satisfied. Suppose that $m_n^\diamond \to \infty$ as $n \to \infty$ and that there exists an unbounded sequence of integers $(q_n)_{n\geqslant 1}$ and a finite constant L satisfying*

$$\sup_{n\geqslant 1} n^3\exp(-n^{1/6}q^{-1}/100) \leqslant L \quad \textit{and} \quad \sup_{n\geqslant 1} n^4 q_n^{-1}\beta_{q_n+1} \leqslant L. \tag{3.2}$$

If we set $\kappa = 2016$ in the definition (2.5) of the penalty $\widehat{\mathrm{pen}}_m$, then there exist a numerical constant $C > 0$ and an integer n_o such that for all $n \geqslant n_o$

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2\Big) \leqslant C\,\Big\{[\mathfrak{b}^2_{m_n^\diamond}(f) \vee n^{-1}\delta^T_{m_n^\diamond}] + n^{-1}[\Sigma^f \vee L]\Big\} \\ \times\, \tau_\infty^2(1+\sigma_\xi^2+\Gamma_2^f)(1+(\Gamma^f_\infty/\sigma_\xi)^2\mathfrak{B}).$$

Note that the penalty $\widehat{\mathrm{pen}}_m$ used in the last assertion depends indeed only on known quantities and, hence the estimator $\widehat{f}_{\widehat{m}}$ with dimension $\widehat{m}$ as in (2.7) is fully data-driven. It is further interesting to compare its upper risk bound given in Corollary 3.3 with the upper bound derived in Theorem 3.1 assuming independent observations. Both upper bounds coincide up to the multiplicative constants, thereby the discussion below Theorem 3.1 applies also here. It remains to underline that (3.2) in Corollary 3.3 imposes a sufficiently fast decay of the sequence of the mixing coefficients $(\beta_k)_{k\geqslant 1}$. Is it interesting to note that an arithmetically decaying sequence of mixing coefficients $(\beta_k)_{k\geqslant 1}$ satisfies (3.2). To be precise, consider a sequence of integers $(q_n)_{n\geqslant 1}$ satisfying $q_n \sim n^r$, i.e., $(n^{-r}q_n)_{n\geqslant 1}$ is bounded away both from zero and infinity, and assume additionally $\beta_k \sim k^{-s}$. In this situation, the condition (3.2) is satisfied whenever $4 - r < rs$ and

$1/6 > r$. In other words, if the sequence of mixing coefficients $(\beta_k)_{k\geqslant 1}$ is sufficiently fast decaying, that is $s > 4(6+\theta)-1$ for some $\theta > 0$, then the condition (3.2) holds true taking, for example, a sequence $q_n \sim n^{1/(6+\theta)}$.

4 Minimax optimality of the data-driven estimator

4.1 Assumptions and notations

We shall access in this section the accuracy of the estimator $\widehat{f}_{\widehat{m}}$ with dimension $\widehat{m}$ selected as in (2.7) by its maximal integrated mean squared error over a class $\mathcal{F}$ of structural functions, that is, $\sup_{f\in\mathcal{F}} \mathbb{E}\|\widehat{f}_{\widehat{m}} - f\|_Z^2$. The class $\mathcal{F}$ reflects prior information on the structural function, e.g., its level of smoothness. It will be determined by means of a weighted norm in L_Z^2 and, hence will be constructed flexibly enough to characterise, in particular, differentiable functions. Given the orthonormal basis $\{u_j\}_{j=1}^{\infty}$ in L_Z^2 and a strictly positive sequence of weights $\mathfrak{a} = (\mathfrak{a}_j)_{j\geqslant 1}$ we define for $h \in L_Z^2$ the weighted norm $\|h\|_{\mathfrak{a}} := (\sum_{j\in\mathbb{N}} \mathfrak{a}_j^{-1}[h]_j^2)^{1/2}$. Furthermore, we denote by $\mathcal{F}_{\mathfrak{a}}$ and $\mathcal{F}_{\mathfrak{a}}^r$ for a constant $r > 0$, respectively, the completion of L_Z^2 with respect to $\|\cdot\|_{\mathfrak{a}}$ and the ellipsoid $\mathcal{F}_{\mathfrak{a}}^r := \{h \in f_{\mathfrak{a}} : \|h\|_{\mathfrak{a}}^2 \leqslant r^2\}$. Observe that $\mathcal{F}_{\mathfrak{a}}$ is a subspace of L_Z^2 for any non-increasing weight sequence $\mathfrak{a}$. Here and subsequently, we assume that there exist a monotonically non-increasing and strictly positive sequence of weights $\mathfrak{f} := (\mathfrak{f}_j)_{j\geqslant 1}$ tending to zero and a constant $r > 0$ such that the structural f belongs to the ellipsoid $\mathcal{F}_{\mathfrak{f}}^r$ which captures all the prior information about the unknown structural function f. Additionally we specify the mapping properties of the conditional expectation operator T and more precisely, we will impose a restriction on the decay of the sequence $(\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}$ which essentially determines δ^T used in the upper bounds given in Theorem 3.1 and 3.2. Denoting by $\mathcal{T}$ the set of all operator mapping L_Z^2 and L_W^2 and given a strictly positive sequence of weights $\mathfrak{t} = (\mathfrak{t}_j)_{j\geqslant 1}$ and a constant $d \geqslant 1$ we define the subset $\mathcal{T}_{\mathfrak{t}}^d$ of $\mathcal{T}$ by

$$\mathcal{T}_{\mathfrak{t}}^d := \Big\{T \in \mathcal{T} : d^{-2}\|f\|_{\mathfrak{t}}^2 \leqslant \|Tf\|_W^2 \leqslant d^2\|f\|_{\mathfrak{t}}^2, \quad \forall f \in L_Z^2\Big\}. \tag{4.1}$$

We notice that each $T \in \mathcal{T}_{\mathfrak{t}}^d$ is injective with $d^{-2} \leqslant \mathfrak{t}_j\|Tu_j\|_W^2 \leqslant d^2$ for all $j \in \mathbb{N}$. Moreover, the sequence $\mathfrak{s} := (\mathfrak{s}_j)_{j\geqslant 1}$ of singular values of T satisfies $d^{-2} \leqslant \mathfrak{t}_j\mathfrak{s}_j^2 \leqslant d^2$, too. We shall emphasise, if $[\nabla_{\mathfrak{t}}]_{\underline{m}}$ denotes the m-dimensional diagonal matrix with diagonal entries $(\mathfrak{t}_j)_{1\leqslant j\leqslant m}$ then for all $T \in \mathcal{T}_{\mathfrak{t}}^d$ holds $\|[T]_{\underline{m}}[\nabla_{\mathfrak{t}}]_{\underline{m}}^{1/2}\|_s \leqslant d$ which in turn implies $\mathfrak{t}_{(m)} := \max_{1\leqslant j\leqslant m} \mathfrak{t}_j = \|[\nabla_{\mathfrak{t}}]_{\underline{m}}^{1/2}\|_s^2 \leqslant d^2\|[T]_{\underline{m}}^{-1}\|_s^2$ for all $m \in \mathbb{N}$. Notice that the link condition (4.1) involves only the basis $\{u_l\}_{l\geqslant 1}$ in L_Z^2. In what follows, we introduce an alternative but stronger condition, which extends the link condition (4.1). We denote by $\mathcal{T}_{\mathfrak{t}}^{d,D}$ for some $D \geqslant d$ the subset of $\mathcal{T}_{\mathfrak{t}}^{d}$ given by

$$\mathcal{T}_{\mathfrak{t}}^{d,D} = \Big\{ T \in \mathcal{T}_{\mathfrak{t}}^{d} : \sup_{m\in\mathbb{N}} \|[\nabla_{\mathfrak{t}}]_{\underline{m}}^{-1/2}[T]_{\underline{m}}^{-1}\|_s \leqslant D \Big\}. \tag{4.2}$$

Obviously, for all $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$ we have $\|[T]_{\underline{m}}^{-1}\|_s^2 \leqslant \|[\nabla_{\mathfrak{t}}]_{\underline{m}}^{1/2}\|_s^2 D^2 = \mathfrak{t}_{(m)} D^2$ and thus $D^{-2} \leqslant d^{-2} \leqslant \mathfrak{t}_{(m)}^{-1}\|[T]_{\underline{m}}^{-1}\|_s^2 \leqslant D^2$ for all $m \in \mathbb{N}$. In other words, the sequence $\mathfrak{t}$ characterises the decay of the sequence $(\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant1}$ for each $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$. It is important to note, that the extended link condition (4.2) guaranties further the convergence of the theoretical approximation $f_m \in \mathcal{U}_m$ given by $[f_m]_{\underline{m}} := [T]_{\underline{m}}^{-1}[g]_{\underline{m}}$ to the structural function f, that is, $\mathfrak{b}_m^2(f) = o(1)$ as $m \to \infty$. Moreover, assuming in addition $f \in \mathcal{F}_{\mathfrak{f}}^r$ the approximation error satisfies $\mathfrak{f}_m^{-1}\mathfrak{b}_m^2(f) \leqslant 4D^4r^2$ due to Lemma B.9 in the Appendix B. All results of this section are derived under regularity conditions on the structural function f and the conditional expectation operator T described by the sequences $\mathfrak{f}$ and $\mathfrak{t}$, respectively. The next assumption summarises our conditions on these sequences. An illustration is provided below by assuming a "regular decay" of these sequences.

Assumption A.5. *(a) Let $\mathfrak{t} := (\mathfrak{t}_j)_{j\geqslant1}$ be a strictly positive, finite, monotonically non-decreasing sequences of weights with $\mathfrak{t}_1 = 1$.*

(b) Let $\mathfrak{f} := (\mathfrak{f}_j)_{j\geqslant1}$ be strictly positive, monotonically non-increasing sequence of weights with limit zero, $\mathfrak{f}_1 = 1$ and $\|\sum_{j\geqslant1}\mathfrak{f}_j u_j^2\|_\infty \leqslant \tau_{\mathfrak{f},\infty}^2$ for some finite constant $\tau_{\mathfrak{f},\infty} \geqslant 1$.

Note that under Assumption A.5 (a) for each $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$ the matrix $[T]_{\underline{k}}$ is non-singular with $D^{-2} \leqslant \mathfrak{t}^{-1}\|[T]_{\underline{k}}^{-1}\|_s^2 \leqslant D^2$ for all $k \in \mathbb{N}$, and hence the Assumption A.3 (a) holds true. On the other hand side, Assumption A.5 (b) holds in case of a bounded basis $\{u_j\}_{j=1}^\infty$ for any summable weight sequence $\mathfrak{f}$, that is, $\sum_{j\geqslant1}\mathfrak{f}_j < \infty$. More generally, under Assumption A.2 the additional assumption $\sum_{j\geqslant1} j\mathfrak{f}_j < \infty$ is sufficient to ensure Assumption A.5 (b). Furthermore, under Assumption A.5 (b) the elements of $\mathcal{F}_{\mathfrak{f}}^r$ are bounded uniformly, that is, $\|\phi\|_\infty^2 \leqslant \|\sum_{j\geqslant1}\mathfrak{f}_j u_j^2\|_\infty \|\phi\|_{\mathfrak{f}}^2 \leqslant \tau_{\mathfrak{f},\infty}^2 r^2$ for all $\phi \in \mathcal{F}_{\mathfrak{f}}^r$. The last estimate is used in Lemma B.9 in the Appendix B to show that for all $f \in \mathcal{F}_{\mathfrak{f}}^r$ and $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$ the approximation f_m satisfies $\|f - f_m\|_\infty \leqslant 2\tau_{\mathfrak{f},\infty}D^2 r$ and $\|f_m\|_Z^2 \leqslant 4D^4r^2$. Thereby, setting $\Gamma_2^{\mathfrak{f}} := \|\mu\|_{Z,W}^2 \vee 4D^4r^2$ and $\Gamma_2^{\mathfrak{f}} := \|\mu\|_\infty + (1+2D^2)\tau_{\mathfrak{f},\infty} r$ the Assumption A.3 (b) holds with $\Gamma_2^f := \Gamma_2^{\mathfrak{f}}$ and $\Gamma_\infty^f := \Gamma_\infty^{\mathfrak{f}}$ uniformly for all $f \in \mathcal{F}_{\mathfrak{f}}^r$ and $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$.

4.2 Independent observations

A careful inspection of the proof of Theorem 3.1 shows that the constant Σ^f given as in (C.3) can be bounded uniformly by a constant $\Sigma^{\mathfrak{f}}$ as in (E.4) for all $f \in \mathcal{F}_{\mathfrak{f}}^r$. Keep in mind the notation given in (2.3) and (2.4). Let us introduce in analogy to $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}$, $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}$, δ_m^T and Δ_m^T the quantities $\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}-} := \mathrm{M}_n(4D^2\mathfrak{t})$, $\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+} := \mathrm{M}_n(\mathfrak{t}/(4D^2))$, $\delta_m^{\mathfrak{t}} := \delta_m(\mathfrak{t})$ and

$\Delta^{\mathfrak{t}}_m := \Delta_m(\mathfrak{t})$. Under Assumption A.5 it is easily seen that for each $T \in \mathcal{T}^{d,D}_{\mathfrak{t}}$ we have $(1+2\log D)^{-1}D^{-2} \leqslant \delta^T_m/\delta^{\mathfrak{t}}_m \leqslant (1+2\log D)D^2$ for all $m \geqslant 1$ and $\mathrm{M}^{\mathfrak{t}-}_{\mathrm{n}} \leqslant \mathrm{M}^{\mathrm{T}-}_{\mathrm{n}} \leqslant \mathrm{M}^{\mathrm{T}+}_{\mathrm{n}} \leqslant \mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}$ for all $n \geqslant 1$. If we require in addition that $(\log n)(\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1)^2\Delta^{\mathfrak{t}}_{\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1} = o(n)$ as $n \to \infty$, then it holds immediately $(\log n)(\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1)^2\Delta^{T}_{\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1} = o(n)$ as $n \to \infty$. Moreover, the condition is automatically satisfied in both cases considered in the Illustration below.

Theorem 4.1. *Assume an i.i.d. n-sample of (Y, Z, W) obeying (1.1a–1.1b). Let Assumption A.1, A.2 and A.5 be satisfied. Set $\kappa = 144$ in the definition (2.5) of the penalty $\widehat{\mathrm{pen}}_m$. If $T \in \mathcal{T}^{d,D}_{\mathfrak{t}}$, $\log(n)(\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1)^2\Delta^{\mathfrak{t}}_{\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1} = o(n)$ as $n \to \infty$, then there exists a constant $\Sigma^{\mathfrak{f}}$ given as in (E.4) in the Appendix E, and a numerical constant C such that for all $n \geqslant 1$*

$$\sup_{f\in\mathcal{F}^r_{\mathfrak{f}}} \mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|^2_Z\Big) \leqslant C\ \tau^2_\infty D^4(r^2 + \sigma^2_\xi + \Gamma^{\mathfrak{f}}_2)\{\min_{1\leqslant m\leqslant \mathrm{M}^{\mathfrak{t}-}_{\mathrm{n}}}\{[\mathfrak{f}_m \vee \delta^{\mathfrak{t}}_m n^{-1}]\} + n^{-1}\ \Sigma^{\mathfrak{f}}\}.$$

We shall compare the last assertion with the lower bound of the maximal risk over the classes $\mathcal{F}^r_{\mathfrak{f}}$ and $\mathcal{T}^{d,D}_{\mathfrak{t}}$ given, for example, in Johannes and Schwarz [2011] or Chen and Reiß [2011]. Given sequences as in Assumption A.5 let us define

$$m^\diamond_n := \mathop{\arg\min}_{1\leqslant m\leqslant \mathrm{M}^{\mathfrak{t}-}_{\mathrm{n}}}\{[\mathfrak{f}_m \vee n^{-1}\delta^{\mathfrak{t}}_m]\} \quad \text{and} \quad \mathcal{R}^\diamond_n := [\mathfrak{f}_{m^\diamond_n} \vee n^{-1}\delta^{\mathfrak{t}}_{m^\diamond_n}] \tag{4.3}$$

as well as $m^\star_n := \mathop{\arg\min}_{m\geqslant 1}\{[\mathfrak{f}_m \vee n^{-1}\sum_{j=1}^m \mathfrak{t}_j]\}$ and $\mathcal{R}^\star_n := [\mathfrak{f}_{m^\star_n} \vee n^{-1}\sum_{j=1}^{m^\star_n} \mathfrak{t}_j]$. Assuming a sufficiently rich class $\mathcal{P}_\varepsilon$ of error distributions P_ε (c.f. Johannes and Schwarz [2011] or Chen and Reiß [2011] for a precise definition) there exists a constant C such that for all $T \in \mathcal{T}^{d,D}_{\mathfrak{t}}$ we have

$$\inf_{\widetilde{f}}\ \sup_{P_\varepsilon\in\mathcal{P}_\varepsilon}\ \sup_{f\in\mathcal{F}^r_{\mathfrak{f}}} \mathbb{E}\Big(\|\widetilde{f} - f\|^2_Z\Big) \geqslant C\mathcal{R}^\star_n, \quad \text{for all } n \geqslant 1, \tag{4.4}$$

where the infimum is taken over all possible estimators $\widetilde{f}$ of f. Obviously, the fully data-driven estimator $\widehat{f}_{\widehat{m}}$ given in (2.2) attains the lower bound up to a constant if and only if $\mathcal{R}^\diamond_n$ is of the same order as $\mathcal{R}^\star_n$ which leads immediately to the following corollary.

Corollary 4.2. *Let the Assumptions of Theorem 4.1 be satisfied. If $\sup_{n\geqslant 1}\{\mathcal{R}^\diamond_n/\mathcal{R}^\star_n\} < \infty$, then $\sup_{f\in\mathcal{F}^r_{\mathfrak{f}}} \mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|^2_Z\Big) = O(\mathcal{R}^\star_n)$, as $n \to \infty$.*

We shall emphasise that the last assertion establishes the minimax optimality of the fully data-driven estimator $\widehat{f}_{\widehat{m}}$ over the classes $\mathcal{F}^r_{\mathfrak{f}}$ and $\mathcal{T}^{d,D}_{\mathfrak{t}}$. Therefore, the estimator is called adaptive. However, minimax optimality is only attained if the rates $\mathcal{R}^\star_n$ and $\mathcal{R}^\diamond_n$ are of the same order. This is, for example, the case if the following two conditions hold simultaneously true: (i) $m^\star_n \leqslant \mathrm{M}^{\mathfrak{t}-}_{\mathrm{n}}$ and (ii) $\delta^{\mathfrak{t}}_m \leqslant C\sum_{j=1}^m \mathfrak{t}_j$. Considering the Illustration below in case (P-P) (i) and (ii) are satisfied, while in case (P-E) (ii) does not hold true. However, in case (P-E) no loss in terms of the rate occur since the squared bias term dominates the variance term, for a detailed discussion in a deconvolution context, we refer to Butucea and Tsybakov [2007a,b].

Illustration. *We illustrate briefly the last results considering the following two configurations for the sequences* $\mathfrak{f}$ *and* $\mathfrak{t}$ *which are usually studied in the literature (c.f. Hall and Horowitz [2005], Chen and Reiß [2011], Johannes and Schwarz [2011] or Breunig and Johannes [2015]). Let*

(P-P) $\mathfrak{f}_j = j^{-2p}$ *and* $\mathfrak{t}_j = j^{2a}$, $j \geqslant 1$, *with* $p > 1$ *and* $a > 1/2$;

(P-E) $\mathfrak{f}_j = j^{-2p}$ *and* $\mathfrak{t}_j = \exp(j^{2a} - 1)$, $j \geqslant 1$, *with* $p > 1$, $a > 0$;

then Assumption A.5 is satisfied in both cases. Writing for two strictly positive sequences $(\mathfrak{a}_n)_{n\geqslant 1}$ *and* $(\mathfrak{b}_n)_{n\geqslant 1}$ *that* $\mathfrak{a}_n \sim \mathfrak{b}_n$, *if* $(\mathfrak{a}_n/\mathfrak{b}_n)_{n\geqslant 1}$ *is bounded away from* 0 *and infinity, we have*

(P-P) $m_n^\star \sim n^{1/(2p+2a+1)}$ *and* $\mathcal{R}_n^\diamondsuit \sim \mathcal{R}_n^\star \sim n^{-2p/(2p+2a+1)}$;

(P-E) $m_n^\star \sim (\log n - \frac{2p+(2a-1)_+}{2a} \log(\log n))^{1/(2a)}$ *and* $\mathcal{R}_n^\diamondsuit \sim \mathcal{R}_n^\star \sim (\log n)^{-p/a}$.

An increasing value of the parameter a *leads in both cases to a slower rate* $\mathcal{R}_n^\star$, *and hence it is called degree of ill-posedness; cf. Natterer [1984].*

4.3 Dependent observations

We dismiss again the independence assumption and assume weakly dependent observations as introduced in Section 3.2. Moreover, keeping in mind the case of independent observations we replace Assumption A.3 by Assumption A.5 which allows us to derive in (F.3) a constant $\Sigma^{\mathfrak{f}}$ uniformly over the classes $\mathcal{F}_{\mathfrak{f}}^r$ depending amongst others on the quantities $\Gamma_2^{\mathfrak{f}}$, $\Gamma_\infty^{\mathfrak{f}}$ and σ_ξ.

Theorem 4.3. *Assume a sample* $\{(Y_i, Z_i, W_i)\}_{i=1}^n$ *obeying (1.1a–1.1b) where* $\{(Z_i, W_i)\}_{i=1}^n$ *is drawn from a stationary absolutely regular process with mixing coefficients* $(\beta_k)_{k\geqslant 0}$ *satisfying* $\mathfrak{B} := \sum_{k=0}^\infty (k+1)^2\beta_k < \infty$ *and given* $k \geqslant 1$ *set* $\mathfrak{B}_k := \sum_{j=k}^\infty \beta_k \leqslant \mathfrak{B}$. *Let the Assumptions A.1, A.2, A.4 and A.5 be satisfied. Considering the dimension* $m_n^\diamondsuit$ *as in (4.3) let* $k_n := \lfloor (\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2 \Gamma_{ZW} m_n^\diamondsuit \rfloor$ *and* $\kappa_n^f \in [6 + 8(\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2 \mathfrak{B}_{k_n}, 8(1 + (\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2\mathfrak{B})]$. *Set* $\kappa = 288\kappa_n^f$ *in the definition (2.5) of the penalty* $\widehat{\mathrm{pen}}_m$. *If* $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$, $\log(n)(\mathrm{M}_n^{\mathfrak{t}+}+1)^2 \Delta^{\mathfrak{t}}_{\mathrm{M}_n^{\mathfrak{t}+}+1} = o(n)$ *as* $n \to \infty$, *then there exists a constant* $\Sigma^{\mathfrak{f}}$ *given as in (F.3) in the Appendix F, which depends amongst others on* τ_∞, $\Gamma_\infty^{\mathfrak{f}}$, σ_ξ *and* $\mathfrak{B}$, *and a numerical constant* C *such that for all* $1 \leqslant q \leqslant n$

$$\sup_{f\in\mathcal{F}_{\mathfrak{f}}^r} \mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2\Big) \leqslant C\,\Big\{\mathcal{R}_n^\diamondsuit + n^{-1}[\Sigma^{\mathfrak{f}} \vee n^3\exp(-n^{1/6}q^{-1}/100) \vee n^4 q^{-1}\beta_{q_n+1}]\Big\}$$
$$\times\, \tau_\infty^2 D^4(r^2 + \sigma_\xi^2 + \Gamma_2^{\mathfrak{f}})(1 + (\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2\mathfrak{B}).$$

We shall emphasise that the last assertion provides in analogy to Theorem 3.2 a non-asymptotic risk bound for the estimator $\widehat{f}_{\widehat{m}}$ with dimension $\widehat{m}$ as in (2.7) where the quantity κ_n^f used to construct the penalty $\widehat{\text{pen}}_m$ still depends on the mixing coefficients $(\beta_k)_{k\geqslant 0}$. As Corollary 3.3 in Section 3.2 follows directly from Theorem 3.2 the next assertion is an immediate consequence of Theorem 4.3, and hence its proof is omitted.

Corollary 4.4. *Let the assumptions of Theorem 4.3 be satisfied. Suppose that $m_n^\diamond \to \infty$ as $n \to \infty$ and that there exists an unbounded sequence of integers $(q_n)_{n\geqslant 1}$ and a finite constant L satisfying (3.2). If we set $\kappa = 2016$ in the definition (2.5) of the penalty $\widehat{\text{pen}}_m$, then there exist a numerical constant C and an integer n_o such that for all $n \geqslant n_o$*

$$\sup_{f\in\mathcal{F}_{\mathfrak{f}}^r} \mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2\Big) \leqslant C\,\Big\{\mathcal{R}_n^\diamond + n^{-1}[\Sigma^{\mathfrak{f}} \vee L]\Big\} \times \tau_\infty^2 D^4 (r^2 + \sigma_\xi^2 + \Gamma_2^{\mathfrak{f}})(1 + (\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2 \mathfrak{B}).$$

Let us briefly comment on the last result. The additional condition (3.2) is, for example, satisfied if the mixing coefficients β have an arithmetic decay as pointed out below Corollary 3.3. Comparing Corollary 4.4 and Theorem 4.1 we see that both upper bounds coincide up to the multiplicative constants. Keep in mind that exploiting Theorem 4.1 in case of independent observations Corollary 4.2 establishes minimax optimality of the estimator $\widehat{f}_{\widehat{m}}$ with dimension $\widehat{m}$ as in (2.7) whenever the rates $\mathcal{R}_n^\diamond$ and $\mathcal{R}_n^\star$ coincide. Exactly in the same manner from Corollary 4.4 follows the minimax optimality of the estimator $\widehat{f}_{\widehat{m}}$ for weakly mixing observations provided the rates $\mathcal{R}_n^\diamond$ and $\mathcal{R}_n^\star$ coincide. In particular, considering the Illustration in Section 4.2 the estimator $\widehat{f}_{\widehat{m}}$ attains the minimax rates in the mildly and severely ill-posed case (P-P) and (P-E), respectively, without having in advance the knowledge of the case. It remains to underline that the penalty $\widehat{\text{pen}}_m$ used in Corollary 4.4 depends again only on known quantities and, hence the estimator $\widehat{f}_{\widehat{m}}$ with dimension $\widehat{m}$ as in (2.7) is fully data-driven, and thus, adaptive.

Acknowledgements. This work was supported by the IAP research network no. P7/06 of the Belgian Government (Belgian Science Policy), and by the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain".

Appendix: Proofs

A Notations

We begin by defining and recalling notations to be used in all proofs. Given $m \geqslant 1$, $\mathcal{U}_m$ and $\mathcal{V}_m$ denote the subspace of L^2_Z and L^2_W spanned by the functions $\{u_j\}_{j=1}^m$ and $\{v_j\}_{j=1}^m$, respectively. U_m and $\mathrm{U}_m^\perp$ (resp. V_m and $\mathrm{V}_m^\perp$) denote the orthogonal projections on $\mathcal{U}_m$ and its orthogonal complement $\mathcal{U}_m^\perp$, respectively. If K is an operator mapping L^2_Z to L^2_W and if we restrict $\mathrm{V}_m K \mathrm{U}_m$ to an operator from $\mathcal{U}_m$ to $\mathcal{V}_m$, then it can be represented by a matrix $[K]_{\underline{m}}$ with generic entries $\langle v_j, K u_l\rangle_W =: [K]_{j,l}$ for $1 \leqslant j,l \leqslant m$. The spectral norm of $[K]_{\underline{m}}$ is denoted by $\|[K]_{\underline{m}}\|_s$ and the inverse matrix of $[K]_{\underline{m}}$ by $[K]_{\underline{m}}^{-1}$. For $m \geqslant 1$, $\mathrm{Id}_{\underline{m}}$ denotes the m-dimensional identity matrix and for all $x \in \mathbb{R}^m$ we denote by $x^t x =: \|x\|^2$ its the euclidean norm. Furthermore, keeping in mind the notations given in (2.3) and (2.4) we use for all $m \geqslant 1$ and $n \geqslant 1$

$$
\begin{aligned}
&\Delta_m^T = \Delta_m((\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}),\ \Lambda_m^T = \Lambda_m((\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}),\ \delta_m^T = m\Delta_m^T\Lambda_m^T,\\
&\widehat{\Delta}_m = \Delta_m((\|[\widehat{T}]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}),\quad \widehat{\Lambda}_m = \Lambda_m((\|[\widehat{T}]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}),\quad \widehat{\delta}_m = m\widehat{\Delta}_m\widehat{\Lambda}_m,\\
&\widehat{\mathrm{M}} = \mathrm{M}_n\left((\|\widehat{[T]}_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}\right),\quad \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} = \mathrm{M}_n(4(\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}),\quad \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} = \mathrm{M}_n(\tfrac{1}{4}(\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}),\\
&\mathrm{pen}_m = \kappa\sigma_m^2 m\Delta_m^T\Lambda_m^T n^{-1} \quad\text{and}\quad \widehat{\mathrm{pen}}_m = 11\kappa\widehat{\sigma}_m^2 m\widehat{\Delta}_m\widehat{\Lambda}_m n^{-1}.
\end{aligned}
\tag{A.1}
$$

Recall that $\widehat{[T]}_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n [v(W_i)]_{\underline{m}}[u(Z_i)]_{\underline{m}}^t$ and $\widehat{[g]}_{\underline{m}} = \frac{1}{n}\sum_{i=1}^n Y_i[v(Z_i)]_{\underline{m}}$ where $[T]_{\underline{m}} = \mathbb{E}[v(W)]_{\underline{m}}[u(Z)]_{\underline{m}}^t$ and $[g]_{\underline{m}} = \mathbb{E}Y[v(W)]_{\underline{m}}$. Given $f_m := \sum_{j=1}^m [f_m]_j u_j \in \mathcal{U}_m$, $m \geqslant 1$, with $[f_m]_{\underline{m}} = [T]_{\underline{m}}^{-1}[g]_{\underline{m}}$ which is well-defined since $[T]_{\underline{m}}$ is non singular. Let $\xi := \varepsilon - \mu(Z,W)$ with $\mu(Z,W) := \mathbb{E}[\varepsilon|Z,W]$ where $\{\xi_i\}_{i=1}^n$ forms an iid. sample independent of $\{(Z_i,W_i)\}_{i=1}^n$. Given $\Gamma_2^f = \|\mu\|_{Z,W}^2 \vee \|f\|_Z^2 \vee \sup_{m\geqslant 1}\|f_m\|_Z^2$ we note that $\sigma_Y^2 := \mathbb{E}Y^2 \leqslant \sigma_\xi^2 + 2\Gamma_2^f$ and $\sigma_m^2 = 2\{\sigma_Y^2 + \max_{1\leqslant k\leqslant m}\|f_k\|_Z^2\} \leqslant 2\{\sigma_\xi^2 + 3\Gamma_2^f\}$ where $\sigma_m^2 \geqslant \mathbb{E}(Y - f_m(Z))^2$ and $\sigma_\xi^2 = \inf_{\|h\|_{Z,W}<\infty}\mathbb{E}(\varepsilon - h(Z,W))^2 \leqslant \sigma_Y^2 \wedge \mathbb{E}(Y - f_m(Z))^2$. Furthermore, $\mathbb{E}|Y - f_m(Z)|^{2k} \leqslant 2^{2k-1}\{\mathbb{E}(\xi)^{2k} + (\Gamma_\infty^f)^{2k}\}$ with $\Gamma_\infty^f := \|\mu\|_\infty + \|f\|_\infty + \sup_{m\geqslant 1}\|f - f_m\|_\infty$. Define the random matrix $[\Xi]_{\underline{m}} := \widehat{[T]}_{\underline{m}} - [T]_{\underline{m}}$ and random vectors $[B]_{\underline{m}}$, $[S]_{\underline{m}}$ and $[V]_{\underline{m}} := [B]_{\underline{m}} + [S]_{\underline{m}}$ given by their components

$$
[B]_j := \frac{1}{n}\sum_{i=1}^n \xi_i v_j(W_i),\ [S]_j := \frac{1}{n}\sum_{i=1}^n v_j(W_i)\{\mu(Z_i,W_i) + f(Z_i) - f_m(Z_i)\},\ 1 \leqslant j \leqslant m,
$$

where $\widehat{[g]}_{\underline{m}} - \widehat{[T]}_{\underline{m}}[f_m]_{\underline{m}} = [V]_{\underline{m}}$. Note that $\mathbb{E}[V]_{\underline{m}} = 0$, indeed it holds $\mathbb{E}[B]_{\underline{m}} = 0$ due to the mean independence, i.e., $0 = \mathbb{E}(\varepsilon|W) = \mathbb{E}(\xi|W) + \mathbb{E}(\varepsilon|W) = \mathbb{E}(\xi)$, $\mathbb{E}(\mu(Z,W)) = \mathbb{E}(\mathbb{E}(\mu(Z,W)|W)) = \mathbb{E}(\mathbb{E}(\varepsilon|W)) = 0$ and $\mathbb{E}[S]_{\underline{m}} = [Tf]_{\underline{m}} - [Tf_m]_{\underline{m}} = 0$. Define further

$\widehat{\sigma}_Y^2 := n^{-1}\sum_{i=1}^n Y_i^2$, $\widehat{\sigma}_m^2 = 74\{\widehat{\sigma}_Y^2 + \max_{1\leqslant k\leqslant m}\|\widehat{f}_k\|_Z^2\}$, the events

$$\begin{aligned}
&\Omega_m := \{\|\widehat{[T]}_{\underline{m}}^{-1}\|_s \leqslant \sqrt{n}\}, \quad \mho_m := \{4\|[\Xi]_{\underline{m}}\|_s\|[T]_{\underline{m}}^{-1}\|_s \leqslant 1\},\\
&\mathcal{A}_n := \{\sigma_Y^2 \leqslant 2\widehat{\sigma}_Y^2 \leqslant 3\sigma_Y^2\}, \quad \mathcal{B}_n := \{\|[T]_{\underline{k}}^{-1}\|_s\|[\Xi]_{\underline{k}}\|_s \leqslant 1/4, \forall 1 \leqslant k \leqslant (\mathrm{M_n^{T+}}+1)\},\\
&\mathcal{C}_n := \{\|[T]_{\underline{k}}^{-1}[V]_{\underline{k}}\|^2 \leqslant \tfrac{1}{8}(\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2 + \sigma_Y^2), \forall 1 \leqslant k \leqslant \mathrm{M_n^{T+}}\},\\
&\mathcal{E}_n := \Big\{\mathrm{pen}_m \leqslant \widehat{\mathrm{pen}}_m \leqslant 99\,\mathrm{pen}_m; \quad \forall 1 \leqslant m \leqslant \mathrm{M_n^{T+}}\Big\} \cap \Big\{\mathrm{M_n^{T-}} \leqslant \widehat{\mathrm{M}} \leqslant \mathrm{M_n^{T+}}\Big\},
\end{aligned} \tag{A.2}$$

and their complements Ω_m^c, $\mho_m^c$, $\mathcal{A}_n^c$, $\mathcal{B}_n^c$, $\mathcal{C}_n^c$, and $\mathcal{E}_n^c$, respectively. Furthermore, we will denote by C universal numerical constants and by $C(\cdot)$ constants depending only on the arguments. In both cases, the values of the constants may change from line to line.

B Preliminary results

This section gathers preliminary results. Given independent observations $\{(Z_i, W_i)\}_{i=1}^n$ the first assertion provides our key arguments in order to control the deviations of the data-driven selection procedure. Both inequalities are due to Talagrand [1996], the formulation of the first part can be found for example in Klein and Rio [2005], while the second part is based on equation (5.13) in Corollary 2 in Birgé et al. [1998] and stated in this form for example in Comte and Merlevede [2002].

Lemma B.1. *(Talagrand's inequalities) Let $X_1, \dots, X_n$ be independent random variables and let $\overline{\nu_t} = n^{-1}\sum_{i=1}^n [\nu_t(X_i) - \mathbb{E}(\nu_t(X_i))]$ for ν_t belonging to a countable class $\{\nu_t, t \in \mathcal{T}\}$ of measurable functions. Then, there exists a numerical constant $C > 0$ such that*

$$\mathbb{E}\Big(\sup_{t\in\mathcal{T}}|\overline{\nu_t}|^2 - 6H^2\Big)_+ \leqslant C\left[\frac{v}{n}\exp\left(\frac{-nH^2}{6v}\right) + \frac{h^2}{n^2}\exp\left(\frac{-nH}{100h}\right)\right], \tag{B.1}$$

$$\mathbf{P}\Big(\sup_{t\in\mathcal{T}}|\overline{\nu_t}| \geqslant 2H + \lambda\Big) \leqslant 3\exp\left[-\frac{n}{100}\left(\frac{\lambda^2}{v} \wedge \frac{\lambda}{h}\right)\right], \tag{B.2}$$

for any $\lambda > 0$, where

$$\sup_{t\in\mathcal{T}}\sup_{x\in\mathcal{Z}}|\nu_t(x)| \leqslant h, \qquad \mathbb{E}[\sup_{t\in\mathcal{T}}|\overline{\nu_t}|] \leqslant H, \qquad \sup_{t\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(\nu_t(X_i)) \leqslant v.$$

Lemma B.2 – B.4 gather preliminary results if $\{(Z_i, W_i)\}_{i\in\mathbb{Z}}$ is a stationary absolutely regular process with mixing coefficients $(\beta_k)_{k\geqslant 1}$.

Lemma B.2. *Under Assumption A.2 if $\{(Z_i, W_i)\}_{i\in\mathbb{Z}}$ is a stationary absolutely regular process with mixing coefficients $(\beta_k)_{k\geqslant 0}$, then*

$$\sum_{j,l=1}^m \mathbb{V}\mathrm{ar}(\sum_{i=1}^q u_j(Z_i)v_j(W_i)) \leqslant qm^2\tau_\infty^4[1 + 4\sum_{k=1}^{q-1}\beta_k].$$

Proof of Lemma B.2. Due to Lemma 4.1 in Asin and Johannes [2016] which is a direct consequence of Theorem 2.1 in Viennet [1997] there exists a sequence $(b_k)_{k\geqslant 1}$ of measurable functions $b_k : \mathbb{R} \to [0,1]$ with $\mathbb{E}b_k(Z_0, W_0) = \beta((Z_0, W_0),(Z_k, W_k)) \leqslant \beta_k$ such that for any measurable function h with $\mathbb{E}h^2(Z_0, W_0) < \infty$ and any integer q holds

$$\mathbb{V}\mathrm{ar}(\sum_{i=1}^{q} h(Z_i, W_i)) \leqslant q\mathbb{E}\Big\{h^2(Z_0, W_0)(1 + 4\sum_{k=1}^{q-1} b_k(Z_0, W_0)\}.$$

Setting $h(Z, W) = u_j(Z)v_l(W)$ the last assertion together with Assumption A.2 implies

$$\sum_{j,l=1}^{m} \mathbb{V}\mathrm{ar}(\sum_{i=1}^{q} u_j(Z_i)v_l(W_i)) \leqslant q\sum_{j,l=1}^{m} \mathbb{E}\Big\{u_j^2(Z_i)v_l^2(W_i)(1 + 4\sum_{k=1}^{q-1} b_k(Z_0, W_0)\Big\}$$

$$\leqslant qm^2\tau_\infty^4\mathbb{E}\Big\{(1 + 4\sum_{k=1}^{q-1} b_k(Z_0, W_0)\Big\} \leqslant qm^2\tau_\infty^4\Big\{1 + 4\sum_{k=1}^{q-1}\beta_k\Big\}$$

which shows the assertion, and thus completes the proof. □

The proof of the next assertion follows along the lines of the proof of Theorem 2.1 and Lemma 4.1 in Viennet [1997] and we omit the details.

Lemma B.3. *Let $\{(Z_i, W_i)\}_{i\in\mathbb{Z}}$ be a stationary absolutely regular process with mixing coefficients $(\beta_k)_{k\geqslant 0}$ satisfying $\mathfrak{B} := 2\sum_{k=0}^{\infty}(k+1)\beta_k < \infty$. Then*

$$\mathbb{V}\mathrm{ar}(\sum_{i=1}^{n} h(Z_i, W_i)) \leqslant 4n\Big(\mathbb{E}h^2(Z_0, W_0)\Big)^{1/2}\|h\|_\infty\mathfrak{B}^{1/2}.$$

The next Lemma is a direct consequence of Theorem 2.2 in Viennet [1997] and we omit its proof.

Lemma B.4. *Let $\{(Z_i, W_i)\}_{i\in\mathbb{Z}}$ be a stationary absolutely regular process with mixing coefficients $(\beta_k)_{k\geqslant 0}$ satisfying $\mathfrak{B} := \sum_{k=0}^{\infty}(k+1)^2\beta_k < \infty$. Then there exists a numerical constant $C > 0$ such that*

$$\mathbb{E}|\sum_{i=1}^{n}\{h(Z_i, W_i) - \mathbb{E}h(Z_i, W_i)\}|^4 \leqslant Cn^2\|h\|_\infty^p\mathfrak{B}$$

The next assertion is due to Asin and Johannes [2016] Lemma 4.10, and we omit its proof.

Lemma B.5. *Let $\{(Z_i, W_i)\}_{i\in\mathbb{Z}}$ be a stationary absolutely regular process with mixing coefficients $(\beta_k)_{k\geqslant 0}$. Under the Assumptions A.2 and A.4 we have for any $q \geqslant 1$ and $K \in \{0, \dots, q-1\}$*

$$\sum_{j}^{m} \mathbb{V}\mathrm{ar}(\sum_{i=1}^{q} h(Z_i, W_i)v_j(W_i))$$

$$\leqslant qm\{\tau_\infty^2\|h\|_Z^2 + 2\|h\|_\infty^2[\gamma K/\sqrt{m} + 2\tau_\infty^2\sum_{k=K+1}^{q-1}\beta_k]\}. \quad \text{(B.3)}$$

In the remaining part of this section we gather in Lemma B.6 — B.8 preliminary results linking the different notations introduced in the last section.

Lemma B.6. *For all $n, m \geqslant 1$ we have*

$$\left\{\frac{1}{4} < \frac{\|\widehat{[T]}_{\underline{m}}^{-1}\|_s^2}{\|[T]_{\underline{m}}^{-1}\|_s^2} \leqslant 4, \forall\, 1 \leqslant m \leqslant (\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} + 1)\right\} \subset \left\{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} \leqslant \widehat{\mathrm{M}} \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}\right\}.$$

Proof of Lemma B.6. Let $\widehat{\tau}_m = \|\widehat{[T]}_{\underline{m}}^{-1}\|_s^{-2}$ and recall that $1 \leqslant \widehat{\mathrm{M}} \leqslant \lfloor n^{1/4} \rfloor$ with

$$\left\{\widehat{\mathrm{M}} = M\right\} = \begin{cases} & \left\{\frac{\widehat{\tau}_{M+1}}{(M+1)^2} < \alpha_n^{-1}\right\}, \quad M = 1, \\ \left\{\min\limits_{2 \leqslant m \leqslant M} \frac{\widehat{\tau}_m}{m^2} \geqslant \alpha_n^{-1}\right\} & \cap \left\{\frac{\widehat{\tau}_{M+1}}{(M+1)^2} < \alpha_n^{-1}\right\}, \quad 1 < M < \lfloor n^{1/4} \rfloor, \\ \left\{\min\limits_{2 \leqslant m \leqslant M} \frac{\widehat{\tau}_m}{m^2} \geqslant \alpha_n^{-1}\right\}, & \qquad M = \lfloor n^{1/4} \rfloor. \end{cases}$$

For $\tau_m = \|[T]_{\underline{m}}^{-1}\|_s^{-2}$ we proof below the following two assertions

$$\left\{\widehat{\mathrm{M}} < \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}\right\} \subset \left\{\min_{1 \leqslant m \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}} : \frac{\widehat{\tau}_m}{\tau_m} < \frac{1}{4}\right\}, \tag{B.4}$$

$$\left\{\widehat{\mathrm{M}} > \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}\right\} \subset \left\{\max_{1 \leqslant m \leqslant (\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} + 1)} \frac{\widehat{\tau}_m}{\tau_m} \geqslant 4\right\}. \tag{B.5}$$

Obviously, the assertion of Lemma B.6 follows now by combination of (B.4) and (B.5). Consider (B.4) which is trivial in case $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} = 1$. If $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} > 1$ we have $\min\limits_{1 \leqslant m \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}} \frac{\tau_m}{m^2} \geqslant \frac{4}{\alpha_n}$. By exploiting the last estimate we obtain

$$\left\{\widehat{\mathrm{M}} < \lfloor n^{1/4} \rfloor\right\} \cap \left\{\widehat{\mathrm{M}} < \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}\right\} = \bigcup_{M=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} - 1} \left\{\widehat{\mathrm{M}} = M\right\}$$

$$\subset \bigcup_{M=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} - 1} \left\{\frac{\widehat{\tau}_{M+1}}{(M+1)^2} < \alpha_n^{-1}\right\} = \left\{\min_{2 \leqslant m \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}} \frac{\widehat{\tau}_m}{m^2} < \alpha_n^{-1}\right\} \subset \left\{\min_{1 \leqslant m \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}} \frac{\widehat{\tau}_m}{\tau_m} < 1/4\right\}$$

while trivially $\left\{\widehat{\mathrm{M}} = \lfloor n^{1/4} \rfloor\right\} \cap \left\{\widehat{\mathrm{M}} < \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}\right\} = \emptyset$, which proves (B.4) because $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} \leqslant \lfloor n^{1/4} \rfloor$.

Consider (B.5) which is trivial in case $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} = \lfloor n^{1/4} \rfloor$. If $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} < \lfloor n^{1/4} \rfloor$, then $\frac{\tau_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1}}{(\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1)^2} < \alpha_n^{-1}$, and hence

$$\left\{\widehat{\mathrm{M}} > 1\right\} \cap \left\{\widehat{\mathrm{M}} > \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}\right\} = \bigcup_{M=\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1}^{\lfloor n^{1/4} \rfloor} \left\{\widehat{\mathrm{M}} = M\right\}$$

$$\subset \bigcup_{M=\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1}^{\lfloor n^{1/4} \rfloor} \left\{\min_{2 \leqslant m \leqslant M} \frac{\widehat{\tau}_m}{m^2} \geqslant \alpha_n^{-1}\right\} = \left\{\min_{2 \leqslant m \leqslant (\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1)} \frac{\widehat{\tau}_m}{m^2} \geqslant \alpha_n^{-1}\right\} \subset \left\{\frac{\widehat{\tau}_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1}}{\tau_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1}} \geqslant 4\right\}$$

while trivially $\{\widehat{\mathrm{M}} = 1\} \cap \{\widehat{\mathrm{M}} > \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}\} = \emptyset$ which shows (B.5) and completes the proof. $\square$

Lemma B.7. *Let $\mathcal{A}_n$, $\mathcal{B}_n$ and $\mathcal{C}_n$ as in (A.2). For all $n \geqslant 1$ it holds true that*

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \{\mathrm{pen}_k \leqslant \widehat{\mathrm{pen}}_k \leqslant 99\,\mathrm{pen}_k, 1 \leqslant k \leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}\} \cap \{\mathrm{M}_\mathrm{n}^{\mathrm{T}-} \leqslant \widehat{\mathrm{M}} \leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}\}.$$

Proof of Lemma B.7. Let $(\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1) \geqslant k \geqslant 1$. If $\|[T]_{\underline{k}}^{-1}\|_s \|[\Xi]_{\underline{k}}\|_s \leqslant 1/4$, i.e. on the event $\mathcal{B}_n$, it follows by the usual Neumann series argument that $\|(\mathrm{Id}_{\underline{k}} + [\Xi]_{\underline{k}}[T]_{\underline{k}}^{-1})^{-1} - \mathrm{Id}_{\underline{k}}\|_s \leqslant 1/3$. Thus, using the identity $\widehat{[T]}_{\underline{k}}^{-1} = [T]_{\underline{k}}^{-1} - \widehat{[T]}_{\underline{k}}^{-1}\{(\mathrm{Id}_{\underline{k}} + [\Xi]_{\underline{k}}[T]_{\underline{k}}^{-1})^{-1} - \mathrm{Id}_{\underline{k}}\}$ we conclude

$$\begin{aligned} 2\|[T]_{\underline{k}}^{-1}\|_s \leqslant 3\|\widehat{[T]}_{\underline{k}}^{-1}\|_s \leqslant 4\|[T]_{\underline{k}}^{-1}\|_s \quad \text{and} \\ 2\|[T]_{\underline{k}}^{-1}x\| \leqslant 3\|\widehat{[T]}_{\underline{k}}^{-1}x\| \leqslant 4\|[T]_{\underline{k}}^{-1}x\|, \quad \text{for all } x \in \mathbb{R}^k. \end{aligned} \tag{B.6}$$

Thereby, since $\widehat{[T]}_{\underline{k}}^{-1}([V]_{\underline{k}}) = \widehat{[T]}_{\underline{k}}^{-1}\widehat{[g]}_{\underline{k}} - [T]_{\underline{k}}^{-1}[g]_{\underline{k}}$ we conclude

$$\begin{aligned} &\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2 \leqslant (32/9)\|[T]_{\underline{k}}^{-1}[V]_{\underline{k}}\|^2 + 2\|\widehat{[T]}_{\underline{k}}^{-1}\widehat{[g]}_{\underline{k}}\|^2, \\ &\|\widehat{[T]}_{\underline{k}}^{-1}\widehat{[g]}_{\underline{k}}\|^2 \leqslant (32/9)\|[T]_{\underline{k}}^{-1}[V]_{\underline{k}}\|^2 + 2\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2. \end{aligned}$$

Consequently, on $\mathcal{C}_n$ where $\|[T]_{\underline{k}}^{-1}[V]_{\underline{k}}\|^2 \leqslant \frac{1}{8}(\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2 + \sigma_Y^2)$ it follows that

$$\begin{aligned} &(5/9)(\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2 + \sigma_Y^2) \leqslant \sigma_Y^2 + 2\|\widehat{[T]}_{\underline{k}}^{-1}\widehat{[g]}_{\underline{k}}\|^2, \\ &\|\widehat{[T]}_{\underline{k}}^{-1}\widehat{[g]}_{\underline{k}}\|^2 \leqslant (22/9)\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2 + (4/9)\sigma_Y^2. \end{aligned}$$

and thus on $\mathcal{A}_n$, i.e., $\sigma_Y^2 \leqslant 2\widehat{\sigma}_Y^2 \leqslant 3\sigma_Y^2$ we have

$$\begin{aligned} &(5/9)(\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2 + \sigma_Y^2) \leqslant (3/2)\widehat{\sigma}_Y^2 + 2\|\widehat{[T]}_{\underline{k}}^{-1}\widehat{[g]}_{\underline{k}}\|^2, \\ &\|\widehat{[T]}_{\underline{k}}^{-1}\widehat{[g]}_{\underline{k}}\|^2 + \widehat{\sigma}_Y^2 \leqslant (22/9)\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2 + (10/9)\sigma_Y^2. \end{aligned}$$

Combining the last two inequalities we conclude for all $1 \leqslant k \leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}$

$$(5/18)(\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2 + \sigma_Y^2) \leqslant (\|\widehat{[T]}_{\underline{k}}^{-1}\widehat{[g]}_{\underline{k}}\|^2 + \widehat{\sigma}_Y^2) \leqslant (22/9)(\|[T]_{\underline{k}}^{-1}[g]_{\underline{k}}\|^2 + \sigma_Y^2).$$

Since on the event $\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n$ the last estimates and (B.6) hold for all $1 \leqslant k \leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}$ it follows

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \Big\{5\sigma_m^2 \leqslant 18\widehat{\sigma}_m^2 \leqslant 44\sigma_m^2 \text{ and } 4\Delta_m^T \leqslant 9\widehat{\Delta}_m \leqslant 16\Delta_m^T,\ \forall 1 \leqslant m \leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}\Big\}.$$

From $\widehat{\Lambda}_m = \max_{1 \leqslant k \leqslant m} \frac{\log(\|\widehat{[T]}_{\underline{k}}^{-1}\|_s^2 \vee (k+2))}{\log(k+2)}$ it is easily seen that $(4/9) \leqslant \widehat{\Delta}_m/\Delta_m^T \leqslant (16/9)$ implies $1/2 \leqslant (1+\log(9/4))^{-1} \leqslant \widehat{\Lambda}_m/\Lambda_m^T \leqslant (1+\log(16/9)) \leqslant 2$. Taking into account the last estimates and the definitions $\mathrm{pen}_m = \kappa\sigma_m^2 m \Delta_m^T \Lambda_m^T n^{-1}$ and $\widehat{\mathrm{pen}}_m = 11\kappa\widehat{\sigma}_m^2 m \widehat{\Delta}_m \widehat{\Lambda}_m n^{-1}$ we obtain

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \Big\{\mathrm{pen}_m \leqslant \widehat{\mathrm{pen}}_m \leqslant 99\,\mathrm{pen}_m,\ \forall 1 \leqslant m \leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}\Big\}. \tag{B.7}$$

On the other hand, by exploiting successively (B.6) and Lemma B.6 we have

$$\mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \subset \left\{ \frac{4}{9} \leqslant \frac{\|\widehat{[T]}_{\underline{m}}^{-1}\|_s^2}{\|[T]_{\underline{m}}^{-1}\|_s^2} \leqslant \frac{9}{4}, \forall 1 \leqslant m \leqslant (\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} + 1) \right\} \subset \left\{ \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} \leqslant \widehat{\mathrm{M}} \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} \right\}. \tag{B.8}$$

From (B.7) and (B.8) follows the assertion of the lemma, which completes the proof. □

LEMMA B.8. *For all $m, n \geqslant 1$ with $\sqrt{n} \geqslant (4/3)\|[T]_{\underline{m}}^{-1}\|_s$ we have $\mho_m \subset \Omega_m$.*

PROOF OF LEMMA B.8. We observe that $\|\widehat{[T]}_{\underline{m}}^{-1}\|_s \leqslant (4/3)\|[T]_{\underline{m}}^{-1}\|_s$ due to the usual Neumann series argument, if $\|[T]_{\underline{m}}^{-1}\|_s \|[\Xi]_{\underline{m}}\|_s \leqslant 1/4$, and consequently $\mho_m \subset \Omega_m$ whenever $\sqrt{n} \geqslant (4/3)\|[T]_{\underline{m}}^{-1}\|_s$, which proves the lemma. □

LEMMA B.9. *Let $g = Tf$ and for each $m \in \mathbb{N}$ define $f_m \in \mathcal{U}_m$ with $[f_m]_{\underline{m}} := [T]_{\underline{m}}^{-1}[g]_{\underline{m}}$. Given sequences $\mathfrak{f}$ and $\mathfrak{t}$ satisfying Assumption A.5, for each $f \in \mathcal{F}_{\mathfrak{f}}^r$ and $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$ we obtain*

$$\sup_{m \geqslant 1} \mathfrak{f}_m^{-1}\|f - f_m\|_Z^2 \leqslant 4D^2d^2r^2, \quad \|f - f_m\|_{\mathfrak{f}}^2 \leqslant 4D^2d^2r^2, \quad \|f_m\|_Z^2 \leqslant 4D^2d^2r^2 \tag{B.9}$$

$$\|f - f_m\|_\infty^2 \leqslant 4\tau_{\mathfrak{f},\infty}^2 D^2d^2r^2, \quad \|f\|_\infty^2 \leqslant \tau_{\mathfrak{f},\infty}^2 r^2. \tag{B.10}$$

PROOF OF LEMMA B.9. The proof of (B.9) can be found in Johannes and Schwarz [2011]. Exploiting $\|h\|_\infty^2 \leqslant \|\sum_{j \geqslant 1} \mathfrak{f}_j u_j^2\|_\infty \|h\|_{\mathfrak{f}}^2 \leqslant \tau_{\mathfrak{f},\infty}^2 \|h\|_{\mathfrak{f}}^2$ and (B.9) we obtain (B.10), which completes the proof. □

C Proof of Theorem 3.1

We assume throughout this section that $\{(Y_i, Z_i, W_i)\}_{i=1}^n$ is an independent and identically distributed sample of the random vector (Y, Z, W) obeying the model equations (1.1a–1.1b). We shall prove below the Propositions C.1 and C.2 which are used in the proof of Theorem 3.1. In the proof the propositions we refer to three technical Lemma (C.3 – C.5) which are shown in the end of this section. Moreover, we make use of functions $\Psi^T, \Phi_{1n}, \Phi_{2n}, \Phi_{3n}^T, \Phi_{4n}^T, \Phi_{5n}^T : \mathbb{R}_+ \to \mathbb{R}$ defined by

$$\begin{aligned}
&\Psi^T(x) = \textstyle\sum_{m \geqslant 1} x\|[T]_{\underline{m}}^{-1}\|_s^2 \exp(-m\Lambda_m^T/(6x)),\\
&\Phi_{1n}(x) = xn \exp(-\lfloor n^{1/4} \rfloor \log(n)/(6x)),\\
&\Phi_{2n}(x) = n^{7/6}x^2 \exp(-n^{1/6}/(100x)),\\
&\Phi_{3n}^T(x) = n^3 \exp(-n(\Delta_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}^T)^{-1}/(25600x^2)),\\
&\Phi_{4n}^T(x) = n^3 \exp(-n(\Delta_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1}^T)^{-1}/(6400x))\\
&\Phi_{5n}^T(x) = xn \exp(-\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} \log(n)/(6x)).
\end{aligned} \tag{C.1}$$

We shall emphasise that each function in (C.1) is non decreasing in x and for all $x > 0$, $\Psi^T(x) < \infty$, $\Phi_{1n}(x) = o(1)$ and $\Phi_{2n}(x) = o(1)$ as $n \to \infty$. Moreover, if $\log(n)(\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} + 1)^2 \Delta^T_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1} = o(n)$ as $n \to \infty$ then there exists an integer $n_o := n_o(T, \tau_\infty)$ depending on T and τ_∞ only such that

$$1 \geqslant \sup_{n \geqslant n_o} \Big\{ 1024 \tau_\infty^4 (1 + \Gamma^f_\infty/\sigma_\xi)^2 (\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} + 1)^2 \Delta^T_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}+1} n^{-1} \Big\}, \tag{C.2}$$

and we have also for all $x > 0$, $\Phi^T_{3n}(x) = o(1)$, $\Phi^T_{4n}(x) = o(1)$ and $\Phi^T_{5n}(x) = o(1)$ as $n \to \infty$. Consequently, under Assumption A.1 and A.2 there exists a finite constant Σ^f such that for all $n \geqslant 1$,

$$\begin{aligned}\Sigma^f \geqslant \Big\{ n_o^2 \bigvee n^3 \exp(-n^{1/6}/50) \bigvee \Psi^T\big(1{+}\Gamma^f_\infty/\sigma_\xi\big) \bigvee \Phi_{1n}\big(1{+}\Gamma^f_\infty/\sigma_\xi\big) \bigvee \Phi_{2n}\big(1{+}\Gamma^f_\infty/\sigma_\xi\big) \\ \bigvee \Phi^T_{5n}\big(1 + \Gamma^f_\infty/\sigma_\xi\big) \bigvee \Phi^T_{3n}(1 + \Gamma^f_\infty/\sigma_\xi) \bigvee \Phi^T_{4n}(\|p_{Z,W}\|_\infty) \\ \bigvee \mathbb{E}(\xi/\sigma_\xi)^8 \bigvee (\Gamma^f_\infty/\sigma_\xi)^8 \bigvee (\tau_\infty/\sigma_\xi)^2 \mathbb{E}(\xi/\sigma_\xi)^{12} \Big\}.\end{aligned} \tag{C.3}$$

Proof of Theorem 3.1. We start the proof with the observation that $(\mathrm{pen}_1, \dots, \mathrm{pen}_{\widehat{\mathrm{M}}})$ is by construction a non-decreasing sub-sequence. Therefore, we can apply Lemma 2.1 in Comte and Johannes [2012] which in turn implies for all $1 \leqslant m \leqslant \widehat{\mathrm{M}}$ that

$$\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \leqslant 85[\mathfrak{b}^2_m(f) \vee \widehat{\mathrm{pen}}_m] + 42 \max_{m \leqslant k \leqslant \widehat{\mathrm{M}}} \Big(\|\widehat{f}_k - f_k\|_Z^2 - \widehat{\mathrm{pen}}_k / 6 \Big)_+ \tag{C.4}$$

where $(x)_+ := \max(x, 0)$. Having the last bound in mind we decompose the risk $\mathbb{E}\|\widehat{f}_{\widehat{m}} - f\|_Z^2$ with respect to the event $\mathcal{E}_n$ defined in (A.2) on which the quantities $\widehat{\mathrm{pen}}_m$ and $\widehat{\mathrm{M}}$ are close to their theoretical counterparts pen_m, $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}$ and $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}$ defined in (A.1). To be precise, we consider the elementary identity

$$\mathbb{E}\|\widehat{f}_{\widehat{m}} - f\|_Z^2 = \mathbb{E}\Big(\mathbb{1}_{\mathcal{E}_n}\|\widehat{f}_{\widehat{m}} - f\|_Z^2\Big) + \mathbb{E}\Big(\mathbb{1}_{\mathcal{E}_n^c}\|\widehat{f}_{\widehat{m}} - f\|_Z^2\Big) \tag{C.5}$$

where we bound the two right hand side (rhs) terms separately. The second rhs term we bound with help of Proposition C.2, which leads to

$$\mathbb{E}\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \leqslant \mathbb{E}\Big(\mathbb{1}_{\mathcal{E}_n}\|\widehat{f}_{\widehat{m}} - f\|_Z^2\Big) + C\, n^{-1}\, \tau_\infty^2 (1 + \sigma_\xi^2 + \Gamma^f_2)\Sigma^f. \tag{C.6}$$

Consider the first rhs term. On the event $\mathcal{E}_n$ the upper bound given in (C.4) implies

$$\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \mathbb{1}_{\mathcal{E}_n} \leqslant 582 \min_{1 \leqslant m \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}} \{[\mathfrak{b}^2_m(f) \vee \mathrm{pen}_m]\} + 42 \max_{1 \leqslant k \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}} \Big(\|\widehat{f}_k - f_k\|_Z^2 - \mathrm{pen}_k / 6\Big)_+.$$

Keeping in mind that $\mathrm{pen}_k = 144 \tau_\infty^2 \sigma_k^2 \delta^T_k n^{-1}$ with $\delta^T_k = k \Lambda^T_k \Delta^T_k$ and $\sigma_k^2 \leqslant 2(\sigma_\xi^2 + 3\Gamma^f_2)$ we derive in Proposition C.1 below an upper bound for the expectation of the second rhs term, the remainder term, in the last display. Thereby, we obtain

$$\mathbb{E}\Big(\mathbb{1}_{\mathcal{E}_n}\|\widehat{f}_{\widehat{m}} - f\|_Z^2\Big) \leqslant C\, \tau_\infty^2 (1 + \sigma_\xi^2 + \Gamma^f_2) \{ \min_{1 \leqslant m \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-}} \{[\mathfrak{b}^2_m(f) \vee n^{-1}\delta^T_m]\} + n^{-1}\, \Sigma^f \}.$$

Replacing in (C.6) the first rhs by the last upper bound we obtain the assertion of the theorem, which completes the proof. $\square$

Proposition C.1. *Under the assumptions of Theorem 3.1 there exists a numerical constant C such that for all $n \geqslant 1$*

$$\mathbb{E}\Big\{\max_{1\leqslant k\leqslant \mathrm{M}_n^{\mathrm{T}+}}\Big(\|\widehat{f}_m - f_m\|_Z^2 - 24\tau_\infty^2 m n^{-1}\sigma_m^2\Lambda_m^T\Delta_m^T\Big)_+\Big\} \leqslant Cn^{-1}\tau_\infty^2(1+\sigma_\xi^2+\Gamma_2^f)\Sigma^f.$$

Proof of Proposition C.1. We start the proof with the observation that $\|\widehat{f}_m - f_m\|_Z^2\mathbb{1}_{\Omega_m}\mathbb{1}_{\mho_m} \leqslant 2\|[T]_{\underline{m}}^{-1}\|_s^2\|[V]_{\underline{m}}\|^2$ and $\|\widehat{f}_m - f_m\|_Z^2\mathbb{1}_{\Omega_m}\mathbb{1}_{\mho_m^c} \leqslant n\|[V]_{\underline{m}}\|^2\mathbb{1}_{\mho_m^c}$, and hence

$$\|\widehat{f}_m - f_m\|_Z^2 \leqslant 2\|[T]_{\underline{m}}^{-1}\|_s^2\|[V]_{\underline{m}}\|^2 + n\|[V]_{\underline{m}}\|^2\mathbb{1}_{\mho_m^c} + \|f_m\|_Z^2\mathbb{1}_{\Omega_m^c}.$$

Since $(\Delta_m^T)_{m\geqslant 1}$ as in (A.1) satisfies $\Delta_m^T \geqslant \|[T]_{\underline{m}}^{-1}\|_s^2$ and $\|[V]_{\underline{m}}\|^2\mathbb{1}_{\mho_m^c} \leqslant \|[V]_{\underline{\mathrm{M}_n^{\mathrm{T}+}}}\|^2\sum_{m=1}^{\mathrm{M}_n^{\mathrm{T}+}}\mathbb{1}_{\mho_m^c}$ we obtain

$$\begin{aligned}
&\mathbb{E}\Big\{\max_{1\leqslant m\leqslant \mathrm{M}_n^{\mathrm{T}+}}\Big(\|\widehat{f}_m - f_m\|_Z^2 - 24\tau_\infty^2 m n^{-1}\sigma_m^2\Lambda_m^T\Delta_m^T\Big)_+\Big\}\\
&\qquad\leqslant 2\mathbb{E}\Big\{\max_{1\leqslant m\leqslant \mathrm{M}_n^{\mathrm{T}+}}\|[T]_{\underline{m}}^{-1}\|_s^2\Big(\|[V]_{\underline{m}}\|^2 - 12\tau_\infty^2 m n^{-1}\sigma_m^2\Lambda_m^T\Big)_+\Big\}\\
&\qquad\quad + \mathbb{E}\Big\{n\Big(\|[V]_{\underline{\mathrm{M}_n^{\mathrm{T}+}}}\|^2 - 12\tau_\infty^2\,\mathrm{M}_n^{\mathrm{T}+}\,n^{-1}\sigma_{\mathrm{M}_n^{\mathrm{T}+}}^2\log(n)\Big)_+\Big\}\\
&\qquad\quad + 12\tau_\infty^2\,\mathrm{M}_n^{\mathrm{T}+}\,\sigma_{\mathrm{M}_n^{\mathrm{T}+}}^2\log(n)P\Big(\bigcup_{k=1}^{\mathrm{M}_n^{\mathrm{T}+}}\mho_k^c\Big) + \max_{1\leqslant m\leqslant \mathrm{M}_n^{\mathrm{T}+}}\|f_m\|_Z^2 P\Big(\bigcup_{k=1}^{\mathrm{M}_n^{\mathrm{T}+}}\Omega_k^c\Big)
\end{aligned}\tag{C.7}$$

where we bound separately each of the four rhs terms. In order to bound the first and second rhs term we employ (C.14) in Lemma C.4 with $K = \mathrm{M}_n^{\mathrm{T}+}$ and sequence $\mathfrak{a} = (\mathfrak{a}_m)_{m\geqslant 1}$ given by $\mathfrak{a}_m = \|[T]_{\underline{m}}^{-1}\|_s^2$ and $\mathfrak{a}_m = n\mathbb{1}_{\{m=\mathrm{M}_n^{\mathrm{T}+}\}}$, respectively. Keeping in mind that in both cases $\mathfrak{a}_{(K)}K^2 \leqslant n^{3/2}$ there exists a numerical constant $C > 0$ such that

$$\begin{aligned}
&\mathbb{E}\Big\{\max_{1\leqslant m\leqslant \mathrm{M}_n^{\mathrm{T}+}}\Big(\|\widehat{f}_m - f_m\|_Z^2 - 24\tau_\infty^2 m n^{-1}\sigma_m^2\Lambda_m^T\Delta_m^T\Big)_+\Big\}\\
&\leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Psi^T(1+\Gamma_\infty^f/\sigma_\xi)+\sigma_\xi^2\Phi_{2n}(1+\Gamma_\infty^f/\sigma_\xi)+\sigma_\xi^2\Phi_{5n}^T(1+\Gamma_\infty^f/\sigma_\xi)+\mathbb{E}(\xi/\sigma_\xi)^{12}\Big\}\\
&\qquad + 6\tau_\infty^2\,\mathrm{M}_n^{\mathrm{T}+}\,\sigma_{\mathrm{M}_n^{\mathrm{T}+}}^2\log(n)P\Big(\bigcup_{k=1}^{\mathrm{M}_n^{\mathrm{T}+}}\mho_k^c\Big) + \max_{1\leqslant m\leqslant \mathrm{M}_n^{\mathrm{T}+}}\|f_m\|_Z^2 P\Big(\bigcup_{k=1}^{\mathrm{M}_n^{\mathrm{T}+}}\Omega_k^c\Big)
\end{aligned}$$

with Ψ^T, Φ_{2n} and Φ_{5n}^T as in (C.1), i.e., $\Psi^T(x) = \sum_{m\geqslant 1}x\|[T]_{\underline{m}}^{-1}\|_s^2\exp(-m\Lambda_m^T/(6x))$, $\Phi_{2n}(x) = n^{7/6}x^2\exp(-n^{1/6}/(100x))$ and $\Phi_{5n}^T(x) = xn\exp(-\mathrm{M}_n^{\mathrm{T}+}\log(n)/(6x))$, $x > 0$. Exploiting that $\sigma_m^2 \leqslant 2(\sigma_\xi^2 + 3\Gamma_2^f)$, $\mathrm{M}_n^{\mathrm{T}+}\log(n) \leqslant n$ and $\max_{1\leqslant m\leqslant \mathrm{M}_n^{\mathrm{T}+}}\|f_m\|_Z^2 \leqslant \Gamma_2^f$, replacing the probability $P(\bigcup_{k=1}^{\mathrm{M_n^{T+}}}\Omega_k^c)$ and $P(\bigcup_{k=1}^{\mathrm{M_n^{T+}}}\mho_k^c)$ by its upper bound given in (C.13) and (C.10) in Lemma C.3, respectively, and employing the definition of Σ^f as in (C.3) we obtain the result of the proposition, which completes the proof. □

Proposition C.2. *Under the assumptions of Theorem 3.1 there exists a numerical constant C such that for all $n \geqslant 1$*

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant C\, n^{-1}\, \tau_\infty^2(1+\sigma_\xi^2+\Gamma_2^f)\Sigma^f.$$

Proof of Proposition C.2. We start the proof with the observation that $\|\widehat{f}_m - f_m\|_Z^2\mathbb{1}_{\Omega_m} \leqslant n\|[V]_{\underline{m}}\|^2 \leqslant n\|[V]_{\underline{M}}\|^2$ for all $1 \leqslant m \leqslant M := \lfloor n^{1/4}\rfloor$, and hence $\|\widehat{f}_m - f\|_Z^2\mathbb{1}_{\Omega_m} \leqslant 3n\|[V]_{\underline{M}}\|^2 + 6\Gamma_2^f$ where $\Gamma_2^f \geqslant \|f\|_Z^2 \vee \sup_{m\geqslant 1}\|f_m\|_Z^2$ which together with $\widehat{m} \leqslant M$ implies

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant 3\mathbb{E}\Big\{n\Big(\|[V]_{\underline{M}}\|^2 - 12\tau_\infty^2\sigma_M^2 M\log(n)n^{-1}\Big)_+\Big\} \\ + \{36\tau_\infty^2 M\sigma_M^2\log(n) + 6\Gamma_2^f\}P(\mathcal{E}_n^c) \quad \text{(C.8)}$$

where we bound separately the two rhs terms. In order to bound the first rhs term we employ (C.14) in Lemma C.4 with sequence $\mathfrak{a} = (\mathfrak{a}_m)_{m\geqslant 1}$ given by $\mathfrak{a}_m = n\mathbb{1}_{\{m=K\}}$ and $K = M$ where $K^2\mathfrak{a}_{(K)} \leqslant n^{3/2}$. Thereby, there exists a numerical constant $C > 0$ such that

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Phi_{1n}\Big(1+\Gamma_\infty^f/\sigma_\xi\Big) + \sigma_\xi^2\Phi_{2n}\Big(1+\Gamma_\infty^f/\sigma_\xi\Big) + \mathbb{E}(\xi/\sigma_\xi)^{12}\Big\} \\ + \{36\tau_\infty^2 M\sigma_M^2\log(n) + 6\Gamma_2^f\}P(\mathcal{E}_n^c)$$

with Φ_{1n} and Φ_{2n} as in (C.1), i.e., $\Phi_{1n}(x) = xn\exp(-\lfloor n^{1/4}\rfloor\log(n)/(6x))$ and $\Phi_{2n}(x) := n^{7/6}x^2\exp(-n^{1/6}/(100x))$, $x > 0$. Exploiting further the definition of Σ^f as in (C.3) and that $\sigma_M^2 \leqslant 2\{\sigma_\xi^2 + 3\Gamma_2^f\}$ and $M\log(n) \leqslant n$ the result of the proposition follows now by replacing the probability $P(\mathcal{E}_n^c)$ by its upper bound given in (C.12) in Lemma C.3, which completes the proof. □

Lemma C.3. *Under the assumptions of Theorem 3.1 there exists a numerical constant C such that for all $n \geqslant 1$*

$$\mathbf{P}\Big(\mathcal{A}_n^c\Big) = \mathbf{P}\Big(\{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}^c\Big) \leqslant C\,\Sigma^f\, n^{-2}, \quad \text{(C.9)}$$

$$\mathbf{P}\Big(\mathcal{B}_n^c\Big) = \mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M_n^{T+}}+1}\mho_m^c\Big) \leqslant C\,\Sigma^f\, n^{-2}, \quad \text{(C.10)}$$

$$\mathbf{P}\Big(\mathcal{C}_n^c\Big) \leqslant C\,\Sigma^f\, n^{-2}, \quad \text{(C.11)}$$

$$\mathbf{P}\Big(\mathcal{E}_n^c\Big) \leqslant C\,\Sigma^f\, n^{-2}, \quad \text{(C.12)}$$

$$\mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M_n^{T+}}}\Omega_m\Big) \leqslant C\,\Sigma^f\, n^{-2}. \quad \text{(C.13)}$$

Proof of Lemma C.3. Consider (C.9). Since $Y_1^2/\sigma_Y^2 - 1, \dots, Y_n^2/\sigma_Y^2 - 1$ are independent and and centred random variables with $\mathbb{E}\Big|Y_i^2/\sigma_Y^2 - 1\Big|^4 \leqslant C\sigma_Y^{-8}\mathbb{E}|Y|^8$ it follows from Theorem 2.10 in Petrov [1995] that $\mathbb{E}\Big|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1\Big|^4 \leqslant Cn^{-2}\sigma_Y^{-8}\mathbb{E}|Y|^8$ where $\sigma_Y \geqslant \sigma_\xi$ and $\mathbb{E}|Y|^8 \leqslant C(\mathbb{E}(\xi)^8 + (\Gamma_\infty^f)^8)$ with $\Gamma_\infty^f \geqslant \|\mu\|_\infty \vee \|f\|_\infty$. Employing Markov's inequality and the last bounds we obtain $P\Big(|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1| > 1/2\Big) \leqslant Cn^{-2}(\mathbb{E}(\xi/\sigma_\xi)^8 + (\Gamma_\infty^f/\sigma_\xi)^8)$. Thereby, the assertion (C.9) follows from the last bound by employing the definition of Σ^f given in (C.3) and by exploiting that $\{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}^c \subset \{|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1| > 1/2\}$. Consider (C.10)–(C.12). Let $\mathfrak{a}$ be a sequence given by $\mathfrak{a}_m = \|[T]_{\underline{m}}^{-1}\|_s^2$ where $\mathfrak{a}_{(m)} = \Delta_m^T$ and n_o an integer satisfying C.2, that is, $n \geqslant 1024\tau_\infty^4(1+\Gamma_\infty^f/\sigma_\xi)^2(\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)^2\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1}$ for all $n > n_o$. We distinguish in the following the cases $n \leqslant n_o$ and $n > n_o$. Consider (C.10). Obviously, we have $\mathbf{P}\Big(\mathcal{B}_n^c\Big) \leqslant n^{-2}n_o^2$ for all $1 \leqslant n \leqslant n_o$. On the other hand, given $n \geqslant n_o$ and, hence $n \geqslant 256\tau_\infty^4(\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)^2\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1} = 4c^{-2}\tau_\infty^4K^2\mathfrak{a}_{(K)}$ with sequence $\mathfrak{a} = (\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}$, integer $K = \mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1$ and constant $c = 1/8$ we obtain from (C.23) in Lemma C.5 for all $1 \leqslant m \leqslant (\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)$

$$\mathbf{P}(\mho_m^c) = \mathbf{P}(\|[T]_{\underline{m}}^{-1}\|_s^2\|[\Xi]_m\|_s^2 \geqslant 1/16) \leqslant 3\exp\left[\frac{-n}{6400\|p_{Z,W}\|_\infty\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1}} \vee \frac{-n^{1/2}}{50}\right]$$

and hence, given Φ_{4n}^T as in (C.1) and $\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1 \leqslant n$ it follows

$$\mathbf{P}\Big(\mathcal{B}_n^c\Big) \leqslant (\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)\max_{1\leqslant m\leqslant(\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)}\mathbf{P}(\mho_m^c) \leqslant 3\,n^{-2}\,\Phi_{4n}^T(\|p_{Z,W}\|_\infty)\vee\{n^3\exp(-n^{1/2}/50)\}.$$

By combination of the two cases and employing the definition of Σ^f given in (C.3) we obtain (C.10). The proof of (C.11) follows a long the lines of the proof of (C.10) using (C.15) in Lemma C.4 rather than (C.23) in Lemma C.5. Precisely, if $1 \leqslant n \leqslant n_o$ we have $P(\mathcal{C}_n^c) \leqslant n^{-2}n_o^2$, while given $n > n_o$ and, hence $n \geqslant 1024\tau_\infty^2(1+\Gamma_\infty^f/\sigma_\xi)^2(\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1} = 4c^{-2}\tau_\infty^2(1+\Gamma_\infty^f/\sigma_\xi)^2K\mathfrak{a}_{(K)}$ with sequence $\mathfrak{a}_m = \|[T]_{\underline{m}}^{-1}\|_s^2$, integer $K = \mathrm{M}_\mathrm{n}^{\mathrm{T}+}$ and constant $c = 1/16$ from (C.15) in Lemma C.4 we obtain for all $1 \leqslant m \leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}$

$$\begin{aligned}\mathbf{P}(\|[T]_{\underline{m}}^{-1}[V]_{\underline{m}}\|^2 > \tfrac{1}{8}(\|f_m\|_Z^2 + \sigma_Y^2)) &\leqslant \mathbf{P}(\mathfrak{a}_m\|[V]_{\underline{m}}\|^2 > 16c^2\{2\|f_m\|_Z^2 + 2\sigma_Y^2\}) \\ &\leqslant 3\exp\left[\frac{-n}{25600(1+\Gamma_\infty^f/\sigma_\xi)^2\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}} \vee \frac{-n^{1/6}}{50}\right] + 32(\tau_\infty^2/\sigma_\xi^2)\mathbb{E}(\xi/\sigma_\xi)^{12}\,n^{-3}.\end{aligned}$$

Exploiting the definition of Φ_{3n}^T given in (C.1) implies $\mathbf{P}\Big(\mathcal{C}_n^c\Big) \leqslant 3\{n^3\exp(-n^{1/6}/50)\} \vee \Phi_{3n}^T(1+\Gamma_\infty^f/\sigma_\xi)n^{-2} + 32(\tau_\infty^2/\sigma_\xi^2)\mathbb{E}(\xi/\sigma_\xi)^{12}n^{-2}$ The assertion (C.11) follows employing the definition of Σ^f given in (C.3). Consider (C.12). Due to Lemma B.7 it holds $\mathbf{P}\Big(\mathcal{E}_n^c\Big) \leqslant \mathbf{P}\Big(\mathcal{A}_n^c\Big) + \mathbf{P}\Big(\mathcal{B}_n^c\Big) + \mathbf{P}\Big(\mathcal{C}_n^c\Big)$. Therefore, the assertion (C.12) follows from (C.9)–(C.11). Consider (C.13). We distinguish again the two cases $n \leqslant n_o$ and $n > n_o$, where

$\mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_n^{\mathrm{T}+}}\Omega_m^c\Big) \leqslant n^{-2}n_o^2$ for all $1 \leqslant n \leqslant n_o$. On the other hand, for all $n > n_o$ we have $n \geqslant (16/9)\Delta^T_{\mathrm{M}_n^{\mathrm{T}+}+1} \geqslant (16/9)\|[T]_{\underline{m}}^{-1}\|_s^2$ for all $1 \leqslant m \leqslant \mathrm{M}_n^{\mathrm{T}+}$, and hence from Lemma B.8 follows $\bigcup_{m=1}^{\mathrm{M}_n^{\mathrm{T}+}}\Omega_m^c \subset \bigcup_{m=1}^{\mathrm{M}_n^{\mathrm{T}+}}\mho_m^c \subset \mathcal{B}_n^c$ for all $n > n_o$. Thereby, (C.10) implies (C.13) for all $n > n_o$. By combination of the two cases we obtain (C.13), which completes the proof. □

Lemma C.4. *Given a non negative sequence* $\mathfrak{a} := (\mathfrak{a}_m)_{m\in\mathbb{N}}$ *let* $\Lambda_m^{\mathfrak{a}} := \Lambda_m(\mathfrak{a})$ *as in* (2.3), $\mathfrak{a}_{(K)} := \max_{1\leqslant m\leqslant K}\mathfrak{a}_m$, *for any* $x > 0$, $\Phi_n(x) := n^{7/6}x^2\exp(-n^{1/6}/(100x))$ *and* $\Psi_{\mathfrak{a}}(x) := \sum_{m\geqslant 1}x^2\mathfrak{a}_m\exp(-m\Lambda_m^{\mathfrak{a}}/(6x^2)) < \infty$, *which by construction always exists. If* $\mathfrak{a}_{(K)}K^2 \leqslant n^{3/2}$ *and* $\Gamma_\infty^f \geqslant \sup_{m\geqslant 1}\|\mu + f - f_m\|_\infty < \infty$ *then there exists a numerical constant* C *such that*

$$\mathbb{E}\Big(\max_{1\leqslant m\leqslant K}\mathfrak{a}_m[\|[V]_{\underline{m}}\|^2 - 12\tau_\infty^2\sigma_m^2 m\Lambda_m^{\mathfrak{a}}n^{-1}]\Big)_+ \\ \leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Psi_{\mathfrak{a}}\Big(1+\Gamma_\infty^f/\sigma_\xi\Big) + \sigma_\xi^2\Phi_n\Big(1+\Gamma_\infty^f/\sigma_\xi\Big) + \mathbb{E}(\xi/\sigma_\xi)^{12}\Big\} \quad \text{(C.14)}$$

Moreover, if $n \geqslant 4c^{-2}(1+\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty^2K\mathfrak{a}_{(K)}$ *for* $c > 0$ *then for all* $1 \leqslant m \leqslant K$ *holds*

$$\mathbf{P}\Big(\mathfrak{a}_m\|[V]_{\underline{m}}\|^2 \geqslant 16c^2\{2\sigma_Y^2 + 2\|f_m\|_Z^2\}\Big) \\ \leqslant 3\exp\left[\frac{-nc^2}{100(1+\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{a}_{(K)}} \vee \frac{-n^{1/6}}{50}\right] \\ + (8c^2)^{-1}(\tau_\infty^2/\sigma_\xi^2)\mathbb{E}(\xi/\sigma_\xi)^{12}\, n^{-3}. \quad \text{(C.15)}$$

Proof of Lemma C.4. We intend to apply Talagrand's inequalities given in Lemma B.1 employing the identity $\|[V]_{\underline{m}}\|^2 = \sup_{t\in\mathbb{B}_m}|\overline{\nu_t}|^2$ where $\mathbb{B}_m := \{t \in \mathcal{U}_m : \|t\|_Z \leqslant 1\}$and $\nu_t(\xi, Z, W) = \sum_{j=1}^m(\xi + \mu(Z,W) + f(Z) - f_m(Z))[t]_j v_j(W)$ where $\varepsilon = \xi + \mu(Z,W)$ and ξ and (Z,W) are independent. A direct application, however, is not possible since ξ and hence, ν_t are generally not uniformly bounded in ξ. Therefore, let us introduce $\xi^b := \xi\mathbb{1}\{|\xi|\leqslant\sigma_\xi n^{1/3}\} - \mathbb{E}\xi\mathbb{1}\{|\xi|\leqslant\sigma_\xi n^{1/3}\}$ and $\xi^u := \xi - \xi^b$. Setting $\nu_t^b(\xi, Z, W) := \nu_t(\xi^b, Z, W)$ and $\nu_t^u := \nu_t - \nu_t^b = \sum_{j=1}^m \xi^u[t]_j v_j(W)$ we have obviously $\overline{\nu_t} = \overline{\nu_t}^b + \overline{\nu_t}^u$. Considering first the assertion (C.14) it follows

$$\mathbb{E}\Big(\max_{1\leqslant m\leqslant K}\Big\{\mathfrak{a}_m[\|[V]_{\underline{m}}\|^2 - 12\tau_\infty^2\sigma_m^2 m\Lambda_m^{\mathfrak{a}}n^{-1}]\Big\}\Big)_+ \\ \leqslant 2\mathbb{E}\Bigg(\max_{1\leqslant m\leqslant K}\{\mathfrak{a}_m[\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^b|^2 - 6\tau_\infty^2\sigma_m^2 m\Lambda_m^{\mathfrak{a}}n^{-1}]\}\Bigg)_+ + 2\mathfrak{a}_{(K)}\mathbb{E}(\sup_{t\in\mathbb{B}_K}|\overline{\nu_t}^u|^2) \quad \text{(C.16)}$$

where we bound separately each rhs term. Consider the second rhs term. Keeping in mind that $\mathbb{E}|\xi|^2\mathbb{1}\{|\xi|^2>\eta^2\} \leqslant \eta^{-10}\mathbb{E}(|\xi|^{12})$ for any $\eta > 0$ and $\sigma_\xi^2 = \mathbb{E}|\xi|^2$ by employing successively the independence of the sample, Assumption A.2 and $\mathfrak{a}_{(K)}K \leqslant n^{3/2}$ we obtain

$$\mathfrak{a}_{(K)}\mathbb{E}\sup_{t\in\mathbb{B}_K}|\overline{\nu_t}^u|^2 \leqslant n^{-1}\tau_\infty^2\mathfrak{a}_{(K)}K\mathbb{E}[|\xi|^2\mathbb{1}_{\{|\xi|>\sigma_\xi n^{1/3}\}}] \leqslant \tau_\infty^2 n^{-1}\mathbb{E}(\xi/\sigma_\xi)^{12}. \tag{C.17}$$

The first rhs term of (C.16) we bound employing Talagrand's inequality (B.1) given in Lemma B.1. To this end, we need to compute the quantities h, H and v verifying the three required inequalities. Employing $|\xi^b| \leqslant 2\sigma_\xi n^{1/3}$ and Assumption A.2 we obtain

$$\begin{aligned}\sup_{t\in\mathbb{B}_m}\|\nu_t^b\|_\infty = \sup_{\xi,Z,W}\Big|(\xi^b+\mu(Z,W)+f(Z)-f_m(Z))^2\sum_{j=1}^m v_j^2(W)\Big|^{1/2}\\ \leqslant \{\sigma_\xi n^{1/3}+\|\mu+f-f_m\|_\infty\}\tau_\infty m^{1/2} =: h.\end{aligned} \tag{C.18}$$

Employing in addition the independence of the sample, the independence between ξ and (Z,W) implying $\mathbb{E}(\xi^b+\mu(Z,W)+f(Z)-f_m(Z))^2 \leqslant \mathbb{E}(\xi)^2+\mathbb{E}(\mu(Z,W)+f(Z)-f_m(Z))^2 = \mathbb{E}(Y-f_m(Z))^2 \leqslant \sigma_m^2$, and $\Lambda_m^{\mathfrak{a}} \geqslant 1$ the quantity H is given by

$$\begin{aligned}\mathbb{E}\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^b|^2 \leqslant n^{-1}\mathbb{E}(\xi^b+\mu(Z,W)+f(Z)-f_m(Z))^2\sum_{j=1}^m v_j^2(W) \leqslant \tau_\infty^2\mathbb{E}(Y-f_m(Z))^2 mn^{-1}\\ \leqslant \tau_\infty^2 2(\mathbb{E}(Y)^2+\|f_m\|_Z^2)mn^{-1} \leqslant \tau_\infty^2\sigma_m^2 m\Lambda_m^{\mathfrak{a}}n^{-1} =: H^2.\end{aligned} \tag{C.19}$$

It remains to calculate the third quantity v. Using successively the independence of ξ and (Z,W) and the uniform distribution of W we obtain

$$\begin{aligned}&\sup_{t\in\mathbb{B}_m} n^{-1}\sum_{i=1}^n \mathbb{V}\mathrm{ar}(\nu_t(\xi_i^b,Z_i,W_i) \leqslant \sup_{t\in\mathbb{B}_m}\mathbb{E}\nu_t^2(\xi^b,Z,W)\\ &= \sup_{t\in\mathbb{B}_m}\mathbb{E}(\xi^b)^2\mathbb{E}(\sum_{j=1}^m v_j(W)[t]_j)^2 + \sup_{t\in\mathbb{B}_m}\mathbb{E}([\mu(Z,W)+f(Z)-f_m(Z)]\sum_{j=1}^m v_j(W)[t]_j)^2\\ &\leqslant \sigma_\xi^2+\|\mu+f-f_m\|_\infty^2 =: v.\end{aligned} \tag{C.20}$$

Evaluating (B.1) of Lemma B.1 with h, H, v given by (C.18), (C.19) and (C.20), respectively, and exploiting $\tau_\infty^2\sigma_m^2 \geqslant \sigma_\xi^2$ and $\|\mu+f-f_m\|_\infty \leqslant \Gamma_\infty^f$ it follows

$$\begin{aligned}&\mathbb{E}\Big(\max_{1\leqslant m\leqslant K}\{\mathfrak{a}_m[\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^b}|^2 - 6\tau_\infty^2 mn^{-1}\sigma_m^2\Lambda_m^{\mathfrak{a}}]\}\Big)_+\\ &\leqslant Cn^{-1}\sum_{m=1}^K\mathfrak{a}_m\Big\{\sigma_\xi^2(1+\Gamma_\infty^f/\sigma_\xi)^2\exp\Big(-\frac{m\Lambda_m^a}{6(1+\Gamma_\infty^f/\sigma_\xi)^2}\Big)\\ &\qquad + \tau_\infty^2\sigma_\xi^2(1+(\Gamma_\infty^f/\sigma_\xi))^2 mn^{-2+2/3}\exp\Big(-n^{1/6}/[100(1+\Gamma_\infty^f/\sigma_\xi)]\Big)\Big\}\end{aligned}$$

Since $\mathfrak{a}_{(K)}K^2 \leqslant n^{3/2}$ and exploiting the definition of $\Psi_{\mathfrak{a}}$ and Φ_n we conclude

$$\mathbb{E}\Big(\max_{1\leqslant m\leqslant K}\{\mathfrak{a}_m[\sup_{t\in\mathbb{B}_m}|\overline{\nu_t^b}|^2 - 6\tau_\infty^2 mn^{-1}\sigma_m^2\Lambda_m^{\mathfrak{a}}]\}\Big)_+ \\ \leqslant Cn^{-1}\{\sigma_\xi^2\Psi_{\mathfrak{a}}(1+\Gamma_\infty^f/\sigma_\xi) + \tau_\infty^2\sigma_\xi^2\Phi_n(1+\Gamma_\infty^f/\sigma_\xi)\}.$$

We obtain the assertion (C.14) by replacing in (C.16) the last bound and (C.17). Consider now (C.15). From $\|[V]_{\underline{m}}\| \leqslant \sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^b| + \sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^u|$, $\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^u| \leqslant \sup_{t\in\mathbb{B}_K}|\overline{\nu_t}^u|$ and $\mathfrak{a}_m \leqslant \mathfrak{a}_{(K)}$ follows for all $1\leqslant m\leqslant K$

$$\mathbf{P}\Big(\|[V]_{\underline{m}}\| \geqslant 4c\mathfrak{a}_m^{-1/2}\Big) \leqslant \mathbf{P}\Big(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^b| \geqslant 2c\mathfrak{a}_m^{-1/2}\Big) + \mathbf{P}\Big(\sup_{t\in\mathcal{B}_K}|\overline{\nu_t}^u| \geqslant 2c\mathfrak{a}_{(K)}^{-1/2}\Big) \quad \text{(C.21)}$$

where we bound separately each rhs term. Consider the second rhs term. Applying successively Markov's inequality, $\mathfrak{a}_{(K)}K \leqslant n$ and (C.17) we obtain

$$\mathbf{P}\Big(\sup_{t\in\mathcal{B}_K}|\overline{\nu_t}^u| \geqslant 2c\mathfrak{a}_{(K)}^{-1/2}\Big) \leqslant (2c)^{-2}\mathfrak{a}_{(K)}\mathbb{E}\sup_{t\in\mathcal{B}_K}|\overline{\nu_t}^u|^2 \leqslant (2c)^{-2}\tau_\infty^2 n^{-3}\mathbb{E}(\xi/\sigma_\xi)^{12} \quad \text{(C.22)}$$

The first rhs term of (C.21) we bound employing Talagrand's inequality (B.2) given in Lemma B.1 with h, H, v as in by (C.18)–(C.20), respectively. Thereby, for all $\lambda > 0$ we have

$$\mathbf{P}\Big(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^b| \geqslant 2\{2\sigma_Y^2 + 2\|f_m\|_Z^2\}^{1/2}m^{1/2}\tau_\infty n^{-1/2} + \lambda\Big) \\ \leqslant 3\exp\left[\frac{-n\lambda^2}{100(\sigma_\xi+\Gamma_\infty^f)^2} \vee \frac{-n^{2/3}\lambda}{100(\sigma_\xi+\Gamma_\infty^f)\tau_\infty m^{1/2}}\right].$$

Since $n \geqslant 4c^{-2}(1+\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty^2 K\mathfrak{a}_{(K)} \geqslant 4c^{-2}\tau_\infty^2 m\mathfrak{a}_m$, letting $\lambda := c\{2\sigma_Y^2+2\|f_m\|_Z^2\}^{1/2}\mathfrak{a}_m^{-1/2}$ and using $\{2\sigma_Y^2 + 2\|f_m\|_Z^2\}^{1/2} \geqslant \sigma_\xi$, $\mathfrak{a}_m \leqslant \mathfrak{a}_{(K)}$ and $n^{1/2}c \geqslant 2(1+\Gamma_\infty^f/\sigma_\xi)\tau_\infty K^{1/2}\mathfrak{a}_{(K)}^{1/2}$ we obtain

$$\mathbf{P}\Big(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^b| \geqslant 2c\{2\sigma_Y^2 + 2\|f_m\|_Z^2\}^{1/2}\mathfrak{a}_m^{-1/2}\Big) \\ \leqslant 3\exp\left[\frac{-nc^2}{100(1+\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{a}_{(K)}} \vee \frac{-n^{1/6}}{50}\right]$$

We obtain the assertion (C.15) by replacing in (C.21) the last bound and (C.22), which completes the proof. □

Lemma C.5. *Let $\mathfrak{a}$ be a non negative sequence and $\mathfrak{a}_{(K)} := \max_{1\leqslant m\leqslant K}\mathfrak{a}_m$. If $n \geqslant 4c^{-2}\tau_\infty^4 K^2\mathfrak{a}_{(K)}$ for $c > 0$ then for all $1\leqslant m\leqslant K$ holds*

$$\mathbf{P}\Big(\mathfrak{a}_m\|[\Xi]_{\underline{m}}\|_s^2 \geqslant 4c^2\Big) \leqslant 3\exp\left[\frac{-nc^2}{100\|p_{Z,W}\|_\infty\mathfrak{a}_{(K)}} \vee \frac{-n^{1/2}}{50}\right] \quad \text{(C.23)}$$

where $p_{Z,W}$ denotes the joint density of Z and W.

Proof of Lemma C.5. We are going to apply Talagrand's inequality (B.2) in Lemma B.1 using $\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t}(x)|^2=\sum_{j,l=1}^m[\Xi]_{j,l}^2\geqslant\|[\Xi]_{\underline{m}}\|_s^2$ where $\nu_t(Z,W)=\sum_{j,l=1}^m[t]_{j,l}u_j(Z)v_l(W)$. Therefore, we compute next the quantities h, H and v verifying the three required inequalities. Exploiting the independence and identical distribution of the sample and Assumption A.2 we obtain

$$\mathbb{E}[\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t}|^2]\leqslant\frac{1}{n}\sum_{j,l=1}^m\mathbb{E}\Big(u_j^2(Z_1)v_l^2(W_1)\Big)\leqslant\frac{m^2}{n}\tau_\infty^4=:H^2,\tag{C.24}$$

$$\sup_{t\in\mathcal{B}_{m^2}}\|\nu_t\|_\infty^2=\sup_{z,w}\sum_{j,l=1}^m u_j^2(z)v_l^2(w)\leqslant m^2\tau_\infty^4=:h^2,\tag{C.25}$$

$$\sup_{t\in\mathcal{B}_{m^2}}\frac{1}{n}\sum_{i=1}^n\mathbb{V}\mathrm{ar}\Big(\nu_t(Z_i,W_i)\Big)\leqslant\|p_{Z,W}\|_\infty\sup_{t\in\mathcal{B}_{m^2}}\|[t]_{\underline{m}}\|^2=\|p_{Z,W}\|_\infty=:v.\tag{C.26}$$

Evaluating (B.2) of Lemma B.1 with h, H, v given by (C.24)–(C.26), respectively, for any $\lambda>0$ we have

$$\mathbf{P}\Big(\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t}|\geqslant 2m\tau_\infty^2n^{-1/2}+\lambda\Big)\leqslant 3\exp\left[\frac{-n\lambda^2}{100\|p_{Z,W}\|_\infty}\vee\frac{-n\lambda}{100m\tau_\infty^2}\right].$$

Since $n\geqslant 4c^{-2}K^2\mathfrak{a}_{(K)}\tau_\infty^4\geqslant 4c^{-2}m^2\mathfrak{a}_m\tau_\infty^4$, $1\leqslant m\leqslant K$, letting $\lambda:=c\mathfrak{a}_m^{-1/2}$ and using $\mathfrak{a}_m\leqslant\mathfrak{a}_{(K)}$ and $n^{1/2}c\geqslant 2K\mathfrak{a}_{(K)}^{1/2}\tau_\infty^2$ we obtain

$$\begin{aligned}\mathbf{P}\left(\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t}|\geqslant 2c\mathfrak{a}_m^{-1/2}\right)&\leqslant 3\exp\left[\frac{-nc^2}{100\|p_{Z,W}\|_\infty\mathfrak{a}_{(K)}}\vee\frac{-nc}{100\tau_\infty^2K\mathfrak{a}_{(K)}^{1/2}}\right]\\&\leqslant 3\exp\left[\frac{-nc^2}{100\|p_{Z,W}\|_\infty\mathfrak{a}_{(K)}}\vee\frac{-n^{1/2}}{50}\right].\end{aligned}$$

A combination of the last bound and $\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t}(x)|\geqslant\|[\Xi]_{\underline{m}}\|_s$ implies the assertion, which completes the proof. □

D Proof of Theorem 3.2

Throughout this section we suppose that $\{(Z_i,W_i)\}_{i\in\mathbb{Z}}$ is a stationary absolutely regular process with mixing coefficients $(\beta_k)_{k\geqslant0}$. The sample $\{(Y_i,Z_i,W_i)\}_{i=1}^n$ still obeys the model (1.1a–1.1b) and the Assumption A.1, i.e., $\{\xi_i:=\varepsilon_i-\mu(Z_i,W_i)\}_{i=1}^n$ forms an iid. sample independent of $\{(Z_i,W_i)\}_{i=1}^n$. We shall prove below the Propositions D.1 and D.2 which are used in the proof of Theorem 3.2. In the proof the propositions we refer to three technical Lemma (D.3 – D.5) which are shown in the end of this section. Moreover, we make use of functions $\Psi^T, \Phi_{1n}, \Phi_{2n}, \Phi^T_{3n}, \Phi^T_{4n}, \Phi^T_{5n} : \mathbb{R}_+ \to \mathbb{R}$ defined by

$$\begin{aligned}
&\Psi^T(x) = \Psi(x) = \textstyle\sum_{m\geqslant 1} x m^{1/2} \|[T]^{-1}_{\underline{m}}\|^2_s \exp(-m^{1/2}\Lambda^T_m/(48x)),\\
&\Phi_{1n}(x) = xn \exp(-\lfloor n^{1/8}\rfloor \log(n)/(48x)),\\
&\Phi_{2n}(x) = n^{7/6}x^2 \exp(-n^{1/6}/(200x)),\\
&\Phi^T_{3n}(x) = n^3 \exp(-n(\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}})^{-1/2}(\Delta^T_{\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}})^{-1}/(204800x)),\\
&\Phi^T_{4n}(x) = n^3 \exp(-n(\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1)^{-1}(\Delta^T_{\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1})^{-1}/(51200x))\\
&\Phi^T_{5n}(x) = xn \exp(-(\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}})^{1/2} \log(n)/(48x)).
\end{aligned} \tag{D.1}$$

We shall emphasise that the functions are non decreasing in x and for all $x > 0$, $\Psi^T(x) < \infty$, $\Phi_{1n}(x) = o(1)$ and $\Phi_{2n}(x) = o(1)$ as $n \to \infty$. Moreover, if $\log(n)(\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1)^2\Delta^T_{\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1} = o(n)$ as $n \to \infty$ then there exists an integer n_o such that

$$1 \geqslant \sup_{n\geqslant n_o}\Big\{1024\tau^4_\infty(6 + 8(\Gamma^f_\infty/\sigma_\xi)^2\mathfrak{B})(\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1)^2\Delta^T_{\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1}n^{-1}\Big\}. \tag{D.2}$$

If in addition $q_n(\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1)(\Delta^T_{\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}+1})^{1/2}(\log n) = o(n^{2/3})$ then we have also for all $x > 0$, $\Phi^T_{3n}(x) = o(1)$, $\Phi^T_{4n}(x) = o(1)$ and $\Phi^T_{5n}(x) = o(1)$ as $n \to \infty$. Consequently, under Assumption A.1 and A.2 there exists a finite constant Σ^f such that for all $n \geqslant 1$,

$$\begin{aligned}
\Sigma^f \geqslant \Big\{& n^2_o \bigvee \Psi^T\Big(1+(\Gamma^f_\infty/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big) \bigvee \Phi_{1n}\Big(1+(\Gamma^f_\infty/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big) \bigvee \Phi_{2n}\Big(1+\Gamma^f_\infty/\sigma_\xi\Big)\\
&\bigvee \Phi^T_{3n}(1+(\Gamma^f_\infty/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}) \bigvee \Phi^T_{4n}(\|p_{Z,W}\|_\infty\mathfrak{B}^{1/2}\tau^2_\infty) \bigvee \Phi^T_{5n}\Big(1+(\Gamma^f_\infty/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big)\\
&\bigvee \mathbb{E}(\xi/\sigma_\xi)^8 \bigvee (\Gamma^f_\infty/\sigma_\xi)^8\mathfrak{B} \bigvee (\tau_\infty/\sigma_\xi)^2\mathbb{E}(\xi/\sigma_\xi)^{12}\Big\}.
\end{aligned} \tag{D.3}$$

Proof of Theorem 3.2. The proof follows line by line the proof of Theorem 3.1. By using Proposition D.2 rather than Proposition C.2 we obtain similar to (C.6) for all $n \geqslant 1$

$$\begin{aligned}
&\mathbb{E}\|\widehat{f}_{\widehat{m}} - f\|^2_Z \leqslant \mathbb{E}\left(\mathbb{1}_{\mathcal{E}_n}\|\widehat{f}_{\widehat{m}} - f\|^2_Z\right)\\
&+ C\, n^{-1}\, \tau^2_\infty\{\sigma^2_\xi + \Gamma^f_2\}(1 + (\Gamma^f_\infty/\sigma_\xi)^2\mathfrak{B})[\Sigma^f \vee n^3 \exp(-n^{1/6}q^{-1}/100) \vee n^4 q^{-1}\beta_{q+1}].
\end{aligned} \tag{D.4}$$

Consider the first rhs term. On the event $\mathcal{E}_n$ defined in (A.2), on which the quantities $\widehat{\mathrm{pen}}_m$ and $\widehat{\mathrm{M}}$ are close to their theoretical counterparts pen_m, $\mathrm{M}^{\mathrm{T}-}_{\mathrm{n}}$ and $\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}$ defined in (A.1), the upper bound given in (C.4) implies

$$\|\widehat{f}_{\widehat{m}} - f\|^2_Z\mathbb{1}_{\mathcal{E}_n} \leqslant 582[\mathfrak{b}^2_{m^\diamond_n}(f) \vee \mathrm{pen}_{m^\diamond_n}] + 42 \max_{m^\diamond_n\leqslant k\leqslant \mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}}\Big(\|\widehat{f}_k - f_k\|^2_Z - \mathrm{pen}_k/6\Big)_+.$$

Keeping in mind that $\mathrm{pen}_k = 288\kappa^f_n\tau^2_\infty\sigma^2_k\delta^T_k n^{-1}$ with $\delta^T_k = k\Lambda^T_k\Delta^T_k$, $\kappa^f_n \leqslant 8(1+(\Gamma^f_\infty/\sigma_\xi)^2\mathfrak{B})$ and $\sigma^2_k \leqslant 2(\sigma^2_\xi + 3\Gamma^f_2)$ we derive in Proposition D.1 below an upper bound for the expectation of the second rhs term, the remainder term, in the last display. Thereby, we obtain

$$\mathbb{E}\Big(\mathbb{1}\mathcal{E}_n\|\widehat{f}_{\widehat{m}}-f\|_Z^2\Big) \leqslant C\ \Big\{[\mathfrak{b}^2_{m_n^\diamond}(f)\vee n^{-1}\delta^T_{m_n^\diamond}]+n^{-1}[\Sigma^f\vee n^3\exp(-n^{1/6}q^{-1}/100)\vee n^4q^{-1}\beta_{q+1}]\Big\} \\ \times\ \tau_\infty^2(1+\sigma_\xi^2+\Gamma_2^f)(1+(\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B}).$$

Replacing in (D.4) the first rhs by the last upper bound we obtain the assertion of the theorem, which completes the proof. □

Proposition D.1. *Under the assumptions of Theorem 3.2 there exists a numerical constant C such that for all $1\leqslant q\leqslant n$*

$$\mathbb{E}\Big\{\max_{m_n^\diamond\leqslant k\leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\Big(\|\widehat{f}_m-f_m\|_Z^2-48\tau_\infty^2\sigma_m^2\kappa_n^f m\Lambda_m^T\Delta_m^T n^{-1}\Big)_+\Big\} \\ \leqslant C\ n^{-1}\ [\Sigma^f\vee n^3\exp(-n^{1/6}q^{-1}/100)\vee n^4q^{-1}\beta_{q+1}]\ \tau_\infty^2\{\sigma_\xi^2+\Gamma_2^f\}(1+(\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B}).$$

Proof of Proposition D.1. The proof follows along the lines of the proof of Proposition C.1. Similarly to (C.7) we have

$$\mathbb{E}\Big\{\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\Big(\|\widehat{f}_m-f_m\|_Z^2-48\tau_\infty^2\kappa_n^f\sigma_m^2 m\Lambda_m^T\Delta_m^T n^{-1}\Big)_+\Big\} \\ \leqslant 2\mathbb{E}\Big\{\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\|[T]_{\underline{m}}^{-1}\|_s^2\Big(\|[V]_{\underline{m}}\|^2-24\tau_\infty^2\kappa_n^f\sigma_m^2 m\Lambda_m^T n^{-1}\Big)_+\Big\} \\ +\mathbb{E}\Big\{n\Big(\|[V]_{\underline{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}}\|^2-24\tau_\infty^2\sigma^2_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\kappa_n^f\,\mathrm{M}_\mathrm{n}^{\mathrm{T}+}\log(n)n^{-1}\Big)_+\Big\} \\ +24\tau_\infty^2\sigma^2_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\kappa_n^f\,\mathrm{M}_\mathrm{n}^{\mathrm{T}+}\log(n)P\Big(\bigcup_{k=1}^{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\mho_k^c\Big)+\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\|f_m\|_Z^2P\Big(\bigcup_{k=1}^{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\Omega_k^c\Big)$$

where we bound separately each of the four rhs terms. Employing (D.12) in Lemma D.4 with $k=m_n^\diamond$, $K=\mathrm{M}_\mathrm{n}^{\mathrm{T}+}$, sequence $a=(a_m)_{m\geqslant1}$ given by $a_m=\|[T]_{\underline{m}}^{-1}\|_s^2$ and $a_m=n\mathbb{1}_{\{m=\mathrm{M}_\mathrm{n}^{\mathrm{T}+}\}}$, respectively. Keeping in mind that in both cases $\mathfrak{a}_{(K)}K^2\leqslant n^{3/2}$ we bound the first and second rhs term as follows

$$\mathbb{E}\Big\{\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\Big(\|\widehat{f}_m-f_m\|_Z^2-48\tau_\infty^2\kappa_n^f\sigma_m^2 m\Lambda_m^T\Delta_m^T n^{-1}\Big)_+\Big\} \\ \leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Psi^T\Big(1+(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big)+\sigma_\xi^2\Phi_{2n}\Big(1+\Gamma_\infty^f/\sigma_\xi\Big) \\ +\sigma_\xi^2\Phi_{5n}^T\Big(1+(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big)+\sigma_\xi^2(1+\Gamma_\infty^f/\sigma_\xi)^2n^{7/3}q^{-1}\beta_{q+1}+\mathbb{E}(\xi/\sigma_\xi)^6\Big\} \\ +24\tau_\infty^2\kappa_n^f\sigma^2_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\,\mathrm{M}_\mathrm{n}^{\mathrm{T}+}\log(n)P\Big(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\mho_k^c\Big)+\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\|f_m\|_Z^2P\Big(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\Omega_k^c\Big)$$

with Ψ^T, Φ_{2n} and Φ_{5n}^T as in (D.1), i.e., $\Psi^T(x)=\sum_{m\geqslant1}xm^{1/2}\|[T]_{\underline{m}}^{-1}\|_s^2\exp(-m^{1/2}\Lambda_m^T/(48x))$, $\Phi_{2n}(x)=n^{7/6}x^2\exp(-n^{1/6}/(200x))$ and $\Phi_{5n}^T(x)=xn\exp(-(\mathrm{M}_\mathrm{n}^{\mathrm{T}+})^{1/2}\log(n)/(48x))$, $x>$

0. Exploiting that $\sigma_m^2[\mathrm{M}_n^{\mathrm{T}+}] \leqslant 2(\sigma_\xi^2+3\Gamma_2^f)$, $\kappa_n^f \leqslant 8(1+(\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B})$, $\mathrm{M}_n^{\mathrm{T}+}\log(n) \leqslant n$ and $\max_{m_n^\diamond \leqslant m \leqslant \mathrm{M}_n^{\mathrm{T}+}} \|f_m\|_Z^2 \leqslant \Gamma_2^f$, replacing the probability $P(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_n^{\mathrm{T}+}} \Omega_k^c)$ and $P(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_n^{\mathrm{T}+}} \mho_k^c)$ by its upper bound given in (D.10) and (D.7) in Lemma D.3, respectively, and employing the definition of Σ^f as in (D.3) we obtain the result of the proposition, which completes the proof. □

Proposition D.2. *Under the assumptions of Theorem 3.2 there exists a numerical constant C such that for all $1 \leqslant q \leqslant n$*

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant C\, n^{-1}[\Sigma^f \vee n^3 \exp(-n^{1/6}q^{-1}/100) \vee n^4 q^{-1}\beta_{q+1}] \\ \times\ \tau_\infty^2\{\sigma_\xi^2 + \Gamma_2^f\}(1 + (\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B}).$$

Proof of Proposition D.2. The proof follows along the lines of the proof of Proposition C.2. As in (C.8) with $M := \lfloor n^{1/4} \rfloor$ and $\kappa_n^f := 8(1 + (\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B})$ we obtain

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant 3\mathbb{E}\Big\{n\Big(\|[V]_{\underline{M}}\|^2 - 24\tau_\infty^2\kappa_n^f\sigma_M^2 M\log(n)n^{-1}\Big)_+\Big\} \\ + \{72\tau_\infty^2 M\kappa_n^f\sigma_M^2\log(n) + 6\Gamma_2^f\}P(\mathcal{E}_n^c). \quad \text{(D.5)}$$

Considering the first rhs term from (D.12) in Lemma D.4 with sequence $\mathfrak{a} = (\mathfrak{a}_m)_{m\geqslant 1}$ given by $\mathfrak{a}_m = n\mathbb{1}_{\{m=K\}}$ and $k = K = M := \lfloor n^{1/4} \rfloor$ where $K^2\mathfrak{a}_{(K)} \leqslant n^{3/2}$ it follows that

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Phi_{1n}\Big(1 + (\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big) + \sigma_\xi^2\Phi_{2n}\Big(1 + \Gamma_\infty^f/\sigma_\xi\Big) \\ + \sigma_\xi^2(1 + \Gamma_\infty^f/\sigma_\xi)^2 n^{7/3}q^{-1}\beta_{q+1} + \mathbb{E}(\xi/\sigma_\xi)^{12}\Big\} \\ + \{72\tau_\infty^2 M\kappa_n^f\sigma_M^2\log(n) + 6\Gamma_2^f\}P(\mathcal{E}_n^c)$$

with Φ_{1n} and Φ_{2n} as in (D.1), i.e., $\Phi_{1n}(x) = xn\exp(-\lfloor n^{1/8}\rfloor\log(n)/(48x))$ and $\Phi_{2n}(x) := n^{7/6}x^2\exp(-n^{1/6}/(100x))$, $x > 0$. Exploiting further the definition of Σ^f as in (D.3) and that $\sigma_M^2 \leqslant 2\{\sigma_\xi^2 + 3\Gamma_2^f\}$, $\kappa_n^f = 8(1 + (\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B})$ and $M\log(n) \leqslant n$ the result of the proposition follows now by replacing the probability $P(\mathcal{E}_n^c)$ by its upper bound given in (D.9) in Lemma D.3, which completes the proof. □

Lemma D.3. *Under the assumptions of Theorem 3.2 there exists a numerical constant*

C *such that for all* $1 \leqslant q \leqslant n$

$$\mathbf{P}\Big(\mathcal{A}_n^c\Big) = \mathbf{P}\Big(\{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}^c\Big) \leqslant C\, \Sigma^f\, n^{-2}, \tag{D.6}$$

$$\mathbf{P}\Big(\mathcal{B}_n^c\Big) = \mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_n^{\mathrm{T}+}+1} \mho_m^c\Big) \leqslant C\, [\Sigma^f \vee n^3 \exp(-n^{1/2}q^{-1}/50) \vee n^4 q^{-1}\beta_{q+1}]\, n^{-2}, \tag{D.7}$$

$$\mathbf{P}\Big(\mathcal{C}_n^c\Big) \leqslant C\, [\Sigma^f \vee n^3 \exp(-n^{1/6}q^{-1}/100) \vee n^4 q^{-1}\beta_{q+1}]\, n^{-2}, \tag{D.8}$$

$$\mathbf{P}\Big(\mathcal{E}_n^c\Big) \leqslant C\, [\Sigma^f \vee n^3 \exp(-n^{1/6}q^{-1}/100) \vee n^4 q^{-1}\beta_{q+1}]\, n^{-2}, \tag{D.9}$$

$$\mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_n^{\mathrm{T}+}} \Omega_m\Big) \leqslant C\, [\Sigma^f \vee n^3 \exp(-n^{1/2}q^{-1}/50) \vee n^4 q^{-1}\beta_{q+1}]\, n^{-2}. \tag{D.10}$$

Proof of Lemma D.3. Consider (D.6). We note that $Y = \xi + \mu(Z, W) + f(Z)$ and $\sigma_Y^2 = \sigma_\xi^2 + \|\mu + f\|_Z^2$. Thereby, setting $\eta := \mu(Z,W) + f(Z)$ with $\sigma_\eta^2 := \mathbb{E}\eta^2 = \|\mu + f\|_{Z,W}^2$, where η and ξ are independent, we obtain

$$P\Big(|n^{-1}\sum_{i=1}^n Y_i^2 - \sigma_Y^2| > \sigma_Y^2/2\Big) \leqslant P\Big(|n^{-1}\sum_{i=1}^n (\xi_i^2 - \sigma_\xi^2)| > \sigma_Y^2/6\Big) \\ + P\Big(|n^{-1}\sum_{i=1}^n (\eta_i^2 - \sigma_\eta^2)| > \sigma_Y^2/6\Big) + P\Big(|n^{-1}\sum_{i=1}^n 2\eta_i\xi_i| > \sigma_Y^2/6\Big) \tag{D.11}$$

and we bound each rhs term separately. Consider the first rhs term. Since $\xi_1^2 - \sigma_\xi^2, \dots, \xi_n^2 - \sigma_\xi^2$ are independent and and centred random variables with $\mathbb{E}(\xi_i^2 - \sigma_\xi^2)^4 \leqslant C\mathbb{E}(\xi)^8$ it follows from Theorem 2.10 in Petrov [1995] that $\mathbb{E}(n^{-1}\sum_{i=1}^n \xi_i^2 - \sigma_\xi^2)^4 \leqslant Cn^{-k}\mathbb{E}(\xi)^8$. Employing Markov's inequality, the last bound and $\sigma_Y^2 \geqslant \sigma_\xi^2$ we have $P\Big(|n^{-1}\sum_{i=1}^n \xi_i^2 - \sigma_\xi^2| > \sigma_Y^2/6\Big) \leqslant Cn^{-2}\mathbb{E}(\xi/\sigma_\xi)^8$. Consider the second rhs term in (D.14). From Lemma B.4 with $h(Z,W) = \{\mu(Z,W) + f(Z)\}^2 = \eta^2$ follows $\mathbb{E}(n^{-1}\sum_{i=1}^n \eta_i^2 - \sigma_\eta^2)^4 \leqslant n^2\|\mu + f\|_\infty^8 \mathfrak{B}$ and hence, applying Markov's inequality, $\sigma_Y^2 \geqslant \sigma_\xi^2$ and $\Gamma_\infty^f \geqslant \|\mu + f\|_\infty$ we have $P\Big(|n^{-1}\sum_{i=1}^n \eta_i^2 - \sigma_\eta^2| > \sigma_Y^2/6\Big) \leqslant Cn^{-2}(\Gamma_\infty^f/\sigma_\xi)^8\mathfrak{B}$. It remains to consider the last rhs term in (D.14). Keeping in mind, that $\{\xi_i\}_{i=1}^n$ are iid. and independent of $\{\eta_i\}_{i=1}^n$ from Theorem 2.10 in Petrov [1995] follows

$$\mathbb{E}|\sum_{i=1}^n \eta_i\xi_i|^4 = \mathbb{E}(\mathbb{E}[|\sum_{i=1}^n \eta_i\xi_i|^4|\eta_1,\dots,\eta_n]) \leqslant n^2\mathbb{E}(\mathbb{E}[|\eta_1\xi_1|^4|\eta_1,\dots,\eta_n]) = n^2\mathbb{E}(\eta)^4\mathbb{E}(\xi)^4 \\ \leqslant (1/2)n^2\{\mathbb{E}(\eta)^8 + \mathbb{E}(\xi)^8\} \leqslant (1/2)n^2\{\|\mu + f\|_\infty^8 + \mathbb{E}(\xi)^8\}$$

and hence, Markov's inequality, $\sigma_Y^2 \geqslant \sigma_\xi^2$ and $\Gamma_\infty^f \geqslant \|\mu + f\|_\infty$ imply together $P\Big(|\sum_{i=1}^n \eta_i\xi_i| > n\sigma_Y^2/12\Big) \leqslant Cn^{-2}\{(\Gamma_\infty^f/\sigma_\xi)^8 + \mathbb{E}(\xi/\sigma_\xi)^8\}$. Replacing in (D.14) each rhs term by its respective bound, we obtain $P\Big(|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1| > 1/2\Big) \leqslant Cn^{-2}\{(\Gamma_\infty^f/\sigma_\xi)^8 + \mathbb{E}(\xi/\sigma_\xi)^8\}$. Thereby, the assertion (D.6) follows from the last bound by employing the definition of Σ^f given in (D.3) and by exploiting that $\{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}^c \subset$

$\{|n^{-1}\sum_{i=1}^n Y_i^2/\sigma_Y^2 - 1| > 1/2\}$. Consider (D.7)–(D.9). Let $\mathfrak{a}$ be a sequence given by $\mathfrak{a}_m = \|[T]_{\underline{m}}^{-1}\|_s^2$ where $\mathfrak{a}_{(m)} = \Delta_m^T$ and n_o an integer satisfying D.2, that is, $n \geqslant 1024\tau_\infty^4(6+8(\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B})(\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)^2\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1}$ for all $n > n_o$. We distinguish in the following the cases $n \leqslant n_o$ and $n > n_o$. Consider (D.7). Obviously, we have $\mathbf{P}\Big(\mathcal{B}_n^c\Big) \leqslant n^{-2}n_o^2$ for all $1 \leqslant n \leqslant n_o$. On the other hand, given $n \geqslant n_o$ and, hence $n \geqslant 512\tau_\infty^4(1+4\mathfrak{B})(\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)^2\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1} = 8c^{-2}\tau_\infty^4(1+4\mathfrak{B})K^2\mathfrak{a}_{(K)}$ with sequence $\mathfrak{a} = (\|[T]_{\underline{m}}^{-1}\|_s^2)_{m\geqslant 1}$, integer $K = \mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1$ and constant $c = 1/8$ we obtain from (D.23) in Lemma D.5 for all $1 \leqslant m \leqslant (\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)$

$$\begin{aligned}\mathbf{P}(\mho_m^c) &= \mathbf{P}(\|[T]_{\underline{m}}^{-1}\|_s^2\|[\Xi]_m\|_s^2 \geqslant 1/16)\\ &\leqslant 6\exp\left[\frac{-n}{51200\tau_\infty^2\|p_{Z,W}\|_\infty\mathfrak{B}^{1/2}(\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1}} \vee \frac{-n^{1/2}}{50q}\right] + nq^{-1}\beta_{q+1}\end{aligned}$$

and hence, given Φ_{4n}^T as in (D.1) and $\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1 \leqslant n$ it follows

$$\begin{aligned}\mathbf{P}\Big(\mathcal{B}_n^c\Big) &\leqslant (\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)\max_{1\leqslant m\leqslant(\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)}\mathbf{P}(\mho_m^c)\\ &\leqslant 6\left\{\Phi_{4n}^T(\tau_\infty^2\|p_{Z,W}\|_\infty\mathfrak{B}^{1/2}) \vee n^3\exp(-n^{1/2}q^{-1}/50) \vee n^4q^{-1}\beta_{q+1}\right\} n^{-2}.\end{aligned}$$

By combination of the two cases and employing the definition of Σ^f given in (D.3) we obtain (D.7). The proof of (D.8) follows a long the lines of the proof of (D.7) using (D.13) in Lemma D.4 rather than (D.23) in Lemma D.5. Precisely, if $1 \leqslant n \leqslant n_o$ we have $P(\mathcal{C}_n^c) \leqslant n^{-2}n_o^2$, while given $n > n_o$ and, hence $n \geqslant 1024\tau_\infty^2[6+8(\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B}](\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1)\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1} = 4c^{-2}\tau_\infty^2 2[3+4(\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B}]K\mathfrak{a}_{(K)}$ with sequence $\mathfrak{a}_m = \|[T]_{\underline{m}}^{-1}\|_s^2$, integer $K = \mathrm{M}_\mathrm{n}^{\mathrm{T}+}$ and constant $c = 1/16$ from (D.13) in Lemma D.4 we obtain for all $1 \leqslant m \leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}$

$$\begin{aligned}\mathbf{P}(\|[T]_{\underline{m}}^{-1}[V]_{\underline{m}}\|^2 > \tfrac{1}{8}(\|f_m\|_Z^2+\sigma_Y^2)) &\leqslant \mathbf{P}(\mathfrak{a}_m\|[V]_{\underline{m}}\|^2 > 16c^2\{2\|f_m\|_Z^2+2\sigma_Y^2\})\\ &\leqslant 6\exp\left[\frac{-n}{51200(1+4(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2})(\mathrm{M}_\mathrm{n}^{\mathrm{T}+})^{1/2}\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}} \vee \frac{-n^{1/6}}{100q}\right]\\ &\qquad + nq^{-1}\beta_{q+1} + 64(\tau_\infty/\sigma_\xi)^2n^{-3}\mathbb{E}(\xi/\sigma_\xi)^{12}\end{aligned}$$

Exploiting the definition of Φ_{3n}^T given in (D.1) implies $\mathbf{P}\Big(\mathcal{C}_n^c\Big) \leqslant 6\{n^3\exp(-n^{1/6}q^{-1}/100)\} \vee \Phi_{3n}^T(1+(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2})n^{-2} + nq^{-1}\beta_{q+1} + 64(\tau_\infty^2/\sigma_\xi^2)\mathbb{E}(\xi/\sigma_\xi)^{12}n^{-2}$ The assertion (D.8) follows employing the definition of Σ^f given in (D.3). Due to Lemma B.7 it holds $\mathbf{P}\Big(\mathcal{E}_n^c\Big) \leqslant \mathbf{P}\Big(\mathcal{A}_n^c\Big) + \mathbf{P}\Big(\mathcal{B}_n^c\Big) + \mathbf{P}\Big(\mathcal{C}_n^c\Big)$. Therefore, the assertion (D.9) follows from (D.6)–(D.8). Consider (D.10). We distinguish again the two cases $n \leqslant n_o$ and $n > n_o$, where $\mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\Omega_m^c\Big) \leqslant n^{-2}n_o^2$ for all $1 \leqslant n \leqslant n_o$. On the other hand, for all $n > n_o$ we have $n \geqslant (16/9)\Delta^T_{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}+1} \geqslant (16/9)\|[T]_{\underline{m}}^{-1}\|_s^2$ for all $1 \leqslant m \leqslant \mathrm{M}_\mathrm{n}^{\mathrm{T}+}$, and hence from Lemma B.8 follows $\bigcup_{m=1}^{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\Omega_m^c \subset \bigcup_{m=1}^{\mathrm{M}_\mathrm{n}^{\mathrm{T}+}}\mho_m^c \subset \mathcal{B}_n^c$ for all $n > n_o$. Thereby, (D.7) implies (D.10) for all $n > n_o$. By combination of the two cases we obtain (D.10), which completes the proof. □

Lemma D.4. *Given a non negative sequence $\mathfrak{a} := (\mathfrak{a}_m)_{m\in\mathbb{N}}$ let $\Lambda_m^{\mathfrak{a}} := \Lambda_m(\mathfrak{a})$ as in (2.3), $\mathfrak{a}_{(m)} := \max_{1\leqslant k\leqslant m}\mathfrak{a}_k$, for any $x > 0$, $\Phi_n(x) := n^{7/6}x^2\exp(-n^{1/6}/(200x))$ and $\Psi_{\mathfrak{a}}(x) := \sum_{m\geqslant 1} xm^{1/2}\mathfrak{a}_m\exp(-m^{1/2}\Lambda_m^{\mathfrak{a}}/(48x)) < \infty$, which by construction always exists. If $\mathfrak{B} \geqslant 2\sum_{k=0}^{\infty}(k+1)\beta_k$, $\sup_{m\geqslant 1}\|\mu + f - f_m\|_\infty \leqslant \Gamma_\infty^f < \infty$, $\mathfrak{a}_{(K)}K^2 \leqslant n^{3/2}$, $M_k := \lfloor(\sigma_\xi/\Gamma_\infty^f)^2\Gamma_{ZW}k\rfloor$ and $\kappa_n^f \geqslant 2[3 + 4(\Gamma_\infty^f/\sigma_\xi)^2\sum_{j>M_k}\beta_j]$ then there exists a finite numerical constant $C > 0$ such that*

$$\begin{aligned}
&\mathbb{E}\Big(\max_{k\leqslant m\leqslant K}\mathfrak{a}_m[\|[V]_{\underline{m}}\|^2 - 24\tau_\infty^2\kappa_n^f\sigma_m^2 m\Lambda_m^{\mathfrak{a}}n^{-1}]\Big)_+\\
&\leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Psi_{\mathfrak{a}}\Big(1 + (\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big) + \sigma_\xi^2\Phi_n\Big(1 + \Gamma_\infty^f/\sigma_\xi\Big)\\
&\qquad + \sigma_\xi^2(1 + \Gamma_\infty^f/\sigma_\xi)^2 n^{17/6}q^{-1}\beta_{q+1} + \mathbb{E}(\xi/\sigma_\xi)^{12}\Big\} \qquad \text{(D.12)}
\end{aligned}$$

Moreover, if $n \geqslant 4c^{-2}\tau_\infty^2 2[3 + 4(\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B}]K\mathfrak{a}_{(K)}$ and $c > 0$ then for all $1 \leqslant m \leqslant K$ holds

$$\begin{aligned}
&\mathbf{P}\Big(\mathfrak{a}_m\|[V]_{\underline{m}}\|^2 \geqslant 16c^2\{\sigma_Y^2 + \|f_m\|_Z^2\}\Big)\\
&\leqslant 6\exp\left[\frac{-nc^2}{200\{1 + 4(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\}K^{1/2}\mathfrak{a}_{(K)}} \vee \frac{-n^{1/6}}{100q}\right]\\
&\qquad + nq^{-1}\beta_{q+1} + (2c)^{-2}(\tau_\infty/\sigma_\xi)^2 n^{-3}\mathbb{E}(\xi/\sigma_\xi)^{12}. \qquad \text{(D.13)}
\end{aligned}$$

Proof of Lemma D.4. The proof follows a long the lines of the proof of Lemma C.4 and, hence recall the decomposition (C.16), where the second rhs term is still bound by (C.17) employing that $\{\xi_i\}_{i=1}^n$ forms an iid. sample independent of the instruments $\{W_i\}_{i=1}^n$. Precisely, we have

$$\begin{aligned}
&\mathbb{E}\Big(\max_{k\leqslant m\leqslant K}\Big\{\mathfrak{a}_m[\|[V]_{\underline{m}}\|^2 - 24\tau_\infty^2\sigma_m^2 m\Lambda_m^{\mathfrak{a}}n^{-1}]\Big\}\Big)_+\\
&\leqslant 2\mathbb{E}\Bigg(\max_{k\leqslant m\leqslant K}\{\mathfrak{a}_m[\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^b|^2 - 12\tau_\infty^2\sigma_m^2 m\Lambda_m^{\mathfrak{a}}n^{-1}]\}\Bigg)_+ + \tau_\infty^2 n^{-1}\mathbb{E}(\xi/\sigma_\xi)^{12}. \qquad \text{(D.14)}
\end{aligned}$$

Therefore, it remains to consider the first rhs term in (D.14). Consider $(Z_i, W_i)_{i\geqslant 1} = (E_l, O_l)_{l\geqslant 1}$ and $(Z_i^\perp, W_i^\perp)_{i\geqslant 1} = (E_l^\perp, O_l^\perp)_{l\geqslant 1}$ obeying the coupling properties **(P1)**, **(P2)** and **(P3)**. Moreover, introduce analogously $(\xi_i^b)_{i\geqslant 1} = (\vec{\xi}_l^{be}, \vec{\xi}_l^{bo})_{l\geqslant 1}$. Setting $\vec{v}_t(\vec{e}, \vec{z}, \vec{w}) = q^{-1}\sum_{i=1}^q v_t(e_i, z_i, w_i)$ for $n = 2pq$ follows

$$\overline{\nu_t}^b = \tfrac{1}{2}\Big\{\tfrac{1}{p}\sum_{l=1}^p\{\vec{v}_t(\vec{\xi}_l^{be}, E_l) - \mathbb{E}\vec{v}_t(\vec{\xi}_l^{be}, E_l)\} + \tfrac{1}{p}\sum_{l=1}^p\{\vec{v}_t(\vec{\xi}_l^{bo}, O_l) - \mathbb{E}\vec{v}_t(\vec{\xi}_l^{bo}, O_l)\}\Big\} =: \tfrac{1}{2}\{\overline{\nu_t}^{be} + \overline{\nu_t}^{bo}\}.$$

Considering the random variables $(Z_i^\perp, W_i^\perp)_{i\geqslant 1} = (E_l^\perp, O_l^\perp)_{l\geqslant 1}$ rather than $(Z_i, W_i)_{i\geqslant 1} = (E_l, O_l)_{l\geqslant 1}$ we introduce in addition $\overline{\nu_t}^{b\perp} = \frac{1}{2}\{\overline{\nu_t}^{be\perp} + \overline{\nu_t}^{bo\perp}\}$. Keeping in mind that $\overline{\nu_t}^{be}$ and $\overline{\nu_t}^{bo}$ (respectively, $\overline{\nu_t}^{be\perp}$ and $\overline{\nu_t}^{bo\perp}$) are identically distributed, the first rhs term in (D.14) is bounded by

$$\mathbb{E}\Bigg(\max_{k\leqslant m\leqslant \mathrm{M}_n^{\mathrm{T}+}}\{\mathfrak{a}_m[\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^b|^2 - 12\tau_\infty^2\kappa_n^f\sigma_m m\Lambda_m^{\mathfrak{a}}n^{-1}]\}\Bigg)_+ \\ \leqslant 2\mathbb{E}\Bigg(\max_{k\leqslant m\leqslant \mathrm{M}_n^{\mathrm{T}+}}\{\mathfrak{a}_m[\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^{be\perp}|^2 - 6\tau_\infty^2\kappa_n^f\sigma_m m\Lambda_m^{\mathfrak{a}}n^{-1}]\}\Bigg)_+ \\ + 2\mathfrak{a}_K\mathbb{E}\Bigg(\sup_{t\in\mathbb{B}_K}|\overline{\nu_t}^{be\perp} - \overline{\nu_t}^{be}|^2\Bigg)_+ \quad \text{(D.15)}$$

where we consider separately each rhs term starting with the second. From $|\overline{\nu_t}^{be\perp} - \overline{\nu_t}^{be}| = |p^{-1}\sum_{l=1}^p\{\vec{v}_t(\vec{\xi}_l^{be}, E_l^\perp) - \vec{v}_t(\vec{\xi}_l^{be}, E_l)\}| \leqslant 2\|\nu_t^b\|_\infty \sum_{l=1}^p \mathbb{1}\{E_l^\perp \neq E_l\}$ and by exploiting the coupling property **(P3)** and (C.18), $n^{3/2} \geqslant K\mathfrak{a}_{(K)}$ and $\|\mu + f - f_K\|_\infty \leqslant \Gamma_\infty^f$ we obtain

$$\mathfrak{a}_{(K)}\mathbb{E}\Bigg(\sup_{t\in\mathbb{B}_K}|\overline{\nu_t}^{be\perp} - \overline{\nu_t}^{be}|^2\Bigg)_+ \leqslant 4\mathfrak{a}_{(K)}\sup_{t\in\mathbb{B}_K}\|\nu_t^b\|_\infty^2 p\beta_{q+1} \\ \leqslant 4\{\sigma_\xi n^{1/3} + \|\mu + f - f_K\|_\infty\}^2\tau_\infty^2 K\mathfrak{a}_{(K)}p\beta_{q+1} \leqslant 2\sigma_\xi^2\tau_\infty^2\{1 + \Gamma_\infty^f/\sigma_\xi\}^2 n^{17/6}q^{-1}\beta_{q+1} \quad \text{(D.16)}$$

Considering the first rhs term in (D.15) we intend to apply Talagrand's inequality (B.1) given in Lemma B.1. The computation of the quantities h, H and v verifying the three required inequalities is very similar to the calculations given in (C.18)–(C.20). Keeping in mind that $\|\nu_t^{be\perp}\|_\infty = \|\nu_t^{be}\|_\infty \leqslant \|\nu_t^b\|_\infty$ from (C.18) follows

$$\sup_{t\in\mathbb{B}_m}\|\nu_t^{be\perp}\|_\infty \leqslant \{\sigma_\xi n^{1/3} + \|\mu + f - f_m\|_\infty\}\tau_\infty m^{1/2} =: h. \quad \text{(D.17)}$$

Making use of the coupling properties **(P1)-(P3)** we observe that $\{(\vec{\xi}_l^{be}, E_l^\perp)\}_{l=1}^p$ are iid., $\vec{\nu}_t(\vec{\xi}_l^{be}, E_l)$ and $\vec{\nu}_t(\vec{\xi}_l^{be}, E_l^\perp)$ are identically distributed, $\{\vec{\xi}_l^{be}\}_{l=1}^p$ and $\{E_l\}_{l=1}^p$ are independent, and $\vec{\xi}_l^{be}$ has iid. components. Consequently, we have

$$\mathbb{E}\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^{be\perp}|^2 = p^{-1}\sum_{j=1}^m \mathbb{V}\mathrm{ar}\Big[q^{-1}\sum_{i=1}^q\{\xi_i^b + \mu(Z_i, W_i) + f(Z_i) - f_m(Z_i)\}v_j(W_i))\Big] \\ \leqslant p^{-1}\{q^{-1}\sigma_\xi^2 m + q^{-2}\sum_{j=1}^m \mathbb{V}\mathrm{ar}\Big[\sum_{i=1}^q(\mu(Z_i, W_i) + f(Z_i) - f_m(Z_i))v_j(W_i)\Big] \quad \text{(D.18)}$$

Considering the second right hand side term, we apply Lemma B.5, and hence given

$M_k := \lfloor (\sigma_\xi/\Gamma^f_\infty)^2 \Gamma_{ZW} k \rfloor$ we have for all $m \geqslant k$

$$\sum_{j=1}^{m} \mathbb{V}\mathrm{ar}\Big[\sum_{i=1}^{q} (\mu(Z_i, W_i) + f(Z_i) - f_m(Z_i)) v_j(W_i)\Big]$$
$$\leqslant qm\{\tau_\infty^2 \|\mu + f - f_m\|^2_{Z,W} + 2\|\mu + f - f_m\|^2_\infty [\Gamma_{ZW} M/\sqrt{m} + 2\tau_\infty^2 \sum_{j=M+1}^{q-1} \beta_j]\}$$
$$\leqslant qm\{\tau_\infty^2 \|\mu + f - f_m\|^2_{Z,W} + \sigma_\xi^2 [2 + 4(\Gamma^f_\infty/\sigma_\xi)^2 \sum_{j=M_k+1}^{q-1} \beta_j]\}$$

Given $\kappa^f_k \geqslant 2[3 + 4(\Gamma^f_\infty/\sigma_\xi)^2 \sum_{j>M_k} \beta_j]$ by combination of the last bound and (D.18) we obtain for all $m \geqslant k$

$$\mathbb{E} \sup_{t \in \mathbb{B}_m} |\overline{\nu_t}^{be\perp}|^2 \leqslant n^{-1} m \tau_\infty^2 \{\|\mu + f - f_m\|^2_{Z,W} + \sigma_\xi^2\} \kappa^f_k \leqslant \tau_\infty^2 \{\sigma_Y^2 + \|f_m\|^2_Z\} \kappa^f_k m n^{-1}$$
$$\leqslant \tau_\infty^2 \sigma_m^2 \kappa^f_k m \Lambda^{\mathfrak{a}}_m n^{-1} =: H^2. \quad \text{(D.19)}$$

Consider finally v. Employing successively **(P3)**, **(P1)** and Lemma B.3 with $\mathfrak{B} \geqslant 2\sum_{k=0}^\infty (k+1)\beta_k$ we have

$$\sup_{t \in \mathcal{B}_m} \tfrac{1}{p} \sum_{l=1}^{p} \mathbb{V}\mathrm{ar}(\vec{\nu_t}(\vec{\xi}^{\,be}_l, E^\perp_l)) = \sup_{t \in \mathcal{B}_m} \mathbb{V}\mathrm{ar}\Big[q^{-1} \sum_{i=1}^{q} \nu_t(\xi^b_i, Z_i, W_i)\Big]$$
$$\leqslant \frac{\sigma_\xi^2}{q} + \sup_{t \in \mathcal{B}_m} \mathbb{V}\mathrm{ar}\Big[q^{-1} \sum_{i=1}^{q} \{\mu(Z, W) + f(Z_i) - f_m(Z_i)\} \sum_{j=1}^{m} [t]_j v_j(W_i)\Big]$$
$$\leqslant \frac{\sigma_\xi^2}{q} + \frac{4}{q} \sup_{t \in \mathcal{B}_m} \{\|\mu + f - f_m\|^2_\infty \{\mathbb{E}(\sum_{j=1}^{m} [t]_j v_j(W_i))^2\}^{1/2} \|\sum_{j=1}^{m} [t]_j v_j\|_\infty \mathfrak{B}^{1/2}\}$$
$$\leqslant \frac{\sigma_\xi^2}{q} + \frac{4}{q} \|\mu + f - f_m\|^2_\infty m^{1/2} \tau_\infty \mathfrak{B}^{1/2}$$
$$\leqslant q^{-1} m^{1/2} \{\sigma_\xi^2 + 4\|\mu + f - f_m\|^2_\infty \tau_\infty \mathfrak{B}^{1/2}\} =: v \quad \text{(D.20)}$$

Evaluating (B.1) of Lemma B.1 with h, H, v given by (D.17), (D.19) and (D.20), respectively, and exploiting $\kappa^f_k \tau_\infty^2 \sigma_m^2 \geqslant \sigma_\xi^2$ and $\|\mu + f - f_m\|_\infty \leqslant \Gamma^f_\infty$ there exists a numerical constant $C > 0$ such that

$$\mathbb{E}\Big(\max_{k \leqslant m \leqslant K} \{\mathfrak{a}_m [\sup_{t \in \mathbb{B}_m} |\overline{\nu_t}^{be\perp}|^2 - 6\tau_\infty^2 \kappa^f_k m \sigma_m^2 \Lambda^{\mathfrak{a}}_m n^{-1}]\}\Big)_+$$
$$\leqslant C n^{-1} \sum_{m=k}^{K} \mathfrak{a}_m \Big\{\sigma_\xi^2 (1 + (\Gamma^f_\infty/\sigma_\xi)^2 \tau_\infty \mathfrak{B}^{1/2}) m^{1/2} \exp\Big(-\frac{m^{1/2} \Lambda^a_m}{48(1 + (\Gamma^f_\infty/\sigma_\xi)^2 \tau_\infty \mathfrak{B}^{1/2})}\Big)$$
$$+ \tau_\infty^2 \sigma_\xi^2 (1 + \Gamma^f_\infty/\sigma_\xi)^2 m n^{-2+2/3} \exp\Big(-n^{1/6}/[200(1 + \Gamma^f_\infty/\sigma_\xi)]\Big)\Big\}$$

Since $\mathfrak{a}_{(K)}(K)^2 \leqslant n^{3/2}$ from the definition of $\Psi_{\mathfrak{a}}$ and Φ_n it follows

$$\mathbb{E}\Big(\max_{k\leqslant m\leqslant K}\{\mathfrak{a}_m[\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^{be\perp}|^2 - 6\tau_\infty^2\kappa_k^f m\sigma_m^2\Lambda_m^{\mathfrak{a}}n^{-1}]\}\Big)_+ \\ \leqslant Cn^{-1}\{\sigma_\xi^2\Psi_{\mathfrak{a}}(1+(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}) + \tau_\infty^2\sigma_\xi^2\Phi_n(1+\Gamma_\infty^f/\sigma_\xi)\}.$$

Replacing in (D.15) the rhs terms by the last bound and (D.16), respectively, we obtain

$$\mathbb{E}\Big(\max_{k\leqslant m\leqslant K}\{\mathfrak{a}_m[\sup_{t\in\mathbb{B}_m}|\overline{\nu_t}^{b}|^2 - 12\tau_\infty^2\kappa_k^f\sigma_m m\Lambda_m^{\mathfrak{a}}n^{-1}]\}\Big)_+ \\ \leqslant Cn^{-1}\{\sigma_\xi^2\Psi_{\mathfrak{a}}(1+(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}) + \tau_\infty^2\sigma_\xi^2\Phi_n(1+\Gamma_\infty^f/\sigma_\xi)\} \\ + 4\sigma_\xi^2\tau_\infty^2(1+\Gamma_\infty^f/\sigma_\xi)^2 n^{17/6}q^{-1}\beta_{q+1}.$$

which together with (D.14) implies the assertion (D.12). Consider now (D.13). Following the proof of (C.15) we make use of the bound (C.21) where as in (C.22) the second rhs term can still be bounded by applying successively Markov's inequality, $\mathfrak{a}_{(K)}K \leqslant n$ and (C.17). Thereby, we obtain for all $1 \leqslant m \leqslant K$

$$\mathbf{P}\Big(\|[V]_{\underline{m}}\| \geqslant 4c\mathfrak{a}_m^{-1/2}\Big) \leqslant \mathbf{P}\Big(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^{b}| \geqslant 2c\mathfrak{a}_m^{-1/2}\Big) + (2c)^{-2}\tau_\infty^2 n^{-3}\mathbb{E}(\xi/\sigma_\xi)^{12}. \quad \text{(D.21)}$$

Considering the first rhs term we make use of the notations $\overline{\nu_t}^{b} =: \frac{1}{2}\{\overline{\nu_t}^{be} + \overline{\nu_t}^{bo}\}$ and $\overline{\nu_t}^{b\perp} = \frac{1}{2}\{\overline{\nu_t}^{be\perp} + \overline{\nu_t}^{bo\perp}\}$ introduced in the proof of (D.12) above, where $\overline{\nu_t}^{be}$ and $\overline{\nu_t}^{bo}$ (respectively, $\overline{\nu_t}^{be\perp}$ and $\overline{\nu_t}^{bo\perp}$) are identically distributed. Thereby, we have

$$\mathbf{P}\Big(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^{b}| \geqslant 2c\mathfrak{a}_m^{-1/2}\Big) \leqslant 2\mathbf{P}\Big(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^{be}| \geqslant 2c\mathfrak{a}_m^{-1/2}\Big) \\ \leqslant 2\Big[\mathbf{P}(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^{be\perp}| \geqslant 2c\mathfrak{a}_m^{-1/2}) + \mathbf{P}(\bigcup_{i=1}^{p} E_l^\perp \neq E_l)\Big] \\ \leqslant 2\Big[\mathbf{P}(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^{be\perp}| \geqslant 2c\mathfrak{a}_m^{-1/2}) + p\beta_{q+1}\Big] \quad \text{(D.22)}$$

where the last inequality follows from the coupling property **(P3)**. The first rhs term in the last display we bound employing Talagrand's inequality (B.2) given in Lemma B.1 with h, H, v as in (D.17)–(D.20), respectively. Thereby, using $\kappa_1^f = 2[3+4(\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B}]$ we have for all $K \geqslant m \geqslant 1$ and for all $\lambda > 0$

$$\mathbf{P}\Big(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^{be\perp}| \geqslant 2\tau_\infty\{\sigma_Y^2 + \|f_m\|_Z^2\}^{1/2}(\kappa_1^f)^{1/2}m^{1/2}n^{-1/2} + \lambda\Big) \\ \leqslant 3\exp\Big[-\frac{p}{100}\Big(\frac{\lambda^2}{q^{-1}m^{1/2}\{\sigma_\xi^2 + 4(\Gamma_\infty^f)^2\tau_\infty\mathfrak{B}^{1/2}\}} \wedge \frac{\lambda}{(\sigma_\xi+\Gamma_\infty^f)\tau_\infty m^{1/2}n^{1/3}}\Big)\Big].$$

Since $n \geqslant 4c^{-2}\tau_\infty^2\kappa_1^f K\mathfrak{a}_{(K)} \geqslant 4c^{-2}\tau_\infty^2\kappa_1^f m\mathfrak{a}_m$ letting $\lambda := c\{\sigma_Y^2 + \|f_m\|_Z^2\}^{1/2}\mathfrak{a}_m^{-1/2}$ and using $\{\sigma_Y^2 + \|f_m\|_Z^2\}^{1/2} \geqslant \sigma_\xi$, $\mathfrak{a}_m \leqslant \mathfrak{a}_{(K)}$ and $n^{1/2}c \geqslant 2(1+\Gamma_\infty^f/\sigma_\xi)\tau_\infty K^{1/2}\mathfrak{a}_{(K)}^{1/2}$ we obtain

$$\begin{aligned}\mathbf{P}\Big(\sup_{t\in\mathcal{B}_m}|\overline{\nu_t}^{be\perp}| \geqslant 2c\{\sigma_Y^2 + \|f_m\|_Z^2\}^{1/2}\mathfrak{a}_m^{-1/2}\Big)\\ \leqslant 3\exp\left[-\frac{p}{100}\left(\frac{c^2\sigma_\xi^2}{q^{-1}K^{1/2}\{\sigma_\xi^2 + 4(\Gamma_\infty^f)^2\tau_\infty\mathfrak{B}^{1/2}\}\mathfrak{a}_{(K)}} \wedge \frac{c\sigma_\xi}{(\sigma_\xi + \Gamma_\infty^f)\tau_\infty K^{1/2}\mathfrak{a}_{(K)}^{1/2}n^{1/3}}\right)\right]\\ \leqslant 3\exp\left[\frac{-nc^2}{200K^{1/2}\{1+4(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\}\mathfrak{a}_{(K)}} \vee \frac{-n^{1/6}}{100q}\right]\end{aligned}$$

We obtain the assertion (D.13) by replacing successively in (D.22) the first rhs term by the last bound, the resulting bound is then used in (D.21) to derive the assertion, which completes the proof. □

Lemma D.5. *Let $\mathfrak{a}$ be a non negative sequence with $\mathfrak{a}_{(m)} := \max_{1\leqslant k\leqslant m}\mathfrak{a}_k$ and $\mathfrak{B} \geqslant 2\sum_{k=0}^\infty (k+1)\beta_k$. If $n \geqslant 8c^{-2}\tau_\infty^4(1+4\mathfrak{B})K^2\mathfrak{a}_{(K)}$ for $c>0$ then for all $1\leqslant m\leqslant K$ holds*

$$\mathbf{P}\left(\mathfrak{a}_m\|[\Xi]_{\underline{m}}\|_s^2 \geqslant 4c^2\right) \leqslant 6\exp\left[\frac{-nc^2}{800\|p_{Z,W}\|_\infty K\mathfrak{a}_{(K)}\tau_\infty^2\mathfrak{B}^{1/2}} \vee \frac{-n^{1/2}}{50q}\right] + nq^{-1}\beta_{q+1} \quad \text{(D.23)}$$

where $p_{Z,W}$ denotes the joint density of Z and W.

Proof of Lemma D.5. The proof follows a long the lines of the proof of Lemma C.5 applying Talagrand's inequality (B.2) in Lemma B.1 using $\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t}(x)|^2 = \sum_{j,l=1}^m [\Xi]_{j,l}^2 \geqslant \|[\Xi]_{\underline{m}}\|_s^2$ where $\nu_t(Z,W) = \sum_{j,l=1}^m [t]_{j,l}u_j(Z)v_l(W)$. Consider $(Z_i,W_i)_{i\geqslant 1} = (E_l,O_l)_{l\geqslant 1}$ and $(Z_i^\perp,W_i^\perp)_{i\geqslant 1} = (E_l^\perp,O_l^\perp)_{l\geqslant 1}$ which satisfy the coupling properties **(P1)**, **(P2)** and **(P3)**. Let $\vec{v}_t(E_k) = \sum_{j,l=1}^m [t]_{j,l}\vec{\psi}_{j,l}(E_k)$ with $\vec{\psi}_{j,l}(E_k) = q^{-1}\sum_{i\in\mathcal{I}_k^e} u_j(Z_i)v_l(W_i)$, then for $n=2pq$ it follows

$$\overline{\nu_t} = \tfrac{1}{2}\Big\{\tfrac{1}{p}\sum_{l=1}^p\{\vec{v}_t(E_l) - \mathbb{E}\vec{v}_t(E_l)\} + \tfrac{1}{p}\sum_{l=1}^p\{\vec{v}_t(O_l) - \mathbb{E}\vec{v}_t(O_l)\}\Big\} =: \tfrac{1}{2}\{\overline{\nu_t^e} + \overline{\nu_t^o}\}.$$

Considering the random variables $(Z_i^\perp,W_i^\perp)_{i\geqslant 1}$ rather than $(Z_i,W_i)_{i\geqslant 1}$ we introduce in addition $\overline{\nu_t}^\perp = \frac{1}{2}\{\overline{\nu_t^e}^\perp + \overline{\nu_t^o}^\perp\}$. Keeping in mind that $\overline{\nu_t^e}$ and $\overline{\nu_t^o}$ (respectively, $\overline{\nu_t^e}^\perp$ and $\overline{\nu_t^o}^\perp$) are identically distributed, we have

$$\begin{aligned}\mathbf{P}(\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t}| \geqslant x) \leqslant 2\mathbf{P}(\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t^e}| \geqslant x) \leqslant 2\Big[\mathbf{P}(\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t^e}^\perp| \geqslant x) + \mathbf{P}(\bigcup_{i=1}^p E_l^\perp \neq E_l)\Big]\\ \leqslant 2\Big[\mathbf{P}(\sup_{t\in\mathcal{B}_{m^2}}|\overline{\nu_t^e}^\perp| \geqslant x) + p\beta_{q+1}\Big] \quad \text{(D.24)}\end{aligned}$$

where the last inequality follows from the coupling property **(P3)**. The first rhs term in the last display we bound by applying Talagrand's inequality (B.2) in Lemma B.1.

Therefore, we compute next the quantities h, H and v verifying the three required inequalities. Consider h. Exploiting Assumption A.2 we have

$$\sup_{t\in\mathcal{B}_{m^2}} \|\vec{v}_t\|_\infty^2 \leqslant \sum_{j,l=1}^m \|\vec{v}_{j,l}\|_\infty^2 \leqslant \sum_{j,l=1}^m \|u_j^2\|_\infty \|v_l^2\|_\infty \leqslant m^2\tau_\infty^4 =: h^2. \tag{D.25}$$

Consider H. Let $\mathcal{B}_q := \sum_{k=1}^{q-1}\beta_k$. Exploiting successively that $\{\vec{\psi}_{j,l}(E_k^\perp)\}_{k=1}^p$ form an iid. sample, $\vec{\psi}_{j,l}(E_1^\perp)$ and $\vec{\psi}_{j,l}(E_1)$ are identically distributed and Lemma B.2 we obtain

$$\begin{aligned}\mathbb{E}\sup_{t\in\mathcal{B}_{m^2}} |\overline{\nu_t^e}^\perp|^2 = p^{-1}\sum_{j,l=1}^m \mathbb{V}\mathrm{ar}(\vec{\psi}_{j,l}(E_1)) = p^{-1}q^{-2}\sum_{j,l=1}^m \mathbb{V}\mathrm{ar}\Big(\sum_{i\in\mathcal{I}_1^e} u_j(Z_i)v_l(W_i)\Big)\\ \leqslant \frac{2m^2\tau_\infty^4}{n}(1+4\mathcal{B}_q) := H^2. \quad \text{(D.26)}\end{aligned}$$

Consider v. From Lemma B.3 with $h(Z,W) = \sum_{j,l=1}^m [t]_{jl}u_j(Z)v_l(W)$ follows

$$\begin{aligned}\sup_{t\in\mathcal{B}_{m^2}} \frac{1}{p}\sum_{l=1}^p \mathbb{V}\mathrm{ar}\left(\vec{\nu}_t(E_l^\perp)\right) \leqslant 4q^{-1}\sup_{t\in\mathcal{B}_{m^2}}\mathbb{E}\Big[\Big(\sum_{j,l=1}^m [t]_{j,l}u_j(Z_1)u_l(W_1)\Big)^2 b(Z_1,W_1)\Big)\Big]\\ \leqslant 4q^{-1}\sup_{t\in\mathcal{B}_{m^2}}\{\mathbb{E}(\sum_{j,l=1}^m [t]_{j,l}u_j(Z_1)v_l(W_1))^2\}^{1/2}\|\sum_{j,l=1}^m [t]_{jl}u_jv_l\|_\infty\{2\sum_{k=0}^\infty (k+1)\beta_k\}^{1/2}\\ \leqslant 4mq^{-1}\tau_\infty^2\|p_{Z,W}\|_\infty \mathfrak{B}^{1/2} =: v. \quad \text{(D.27)}\end{aligned}$$

Evaluating (B.2) of Lemma B.1 with h, H, v given by (D.26)–(D.27), respectively, for any $\lambda > 0$ we have

$$\begin{aligned}\mathbf{P}\Big(\sup_{t\in\mathcal{B}_{m^2}} |\overline{\nu_t^e}^\perp| \geqslant 2\sqrt{2}m\tau_\infty^2(1+4\mathfrak{B}_q)^{1/2}n^{-1/2}+\lambda\Big)\\ \leqslant 3\exp\left[\frac{-n\lambda^2}{800\|p_{Z,W}\|_\infty m\tau_\infty^2\mathfrak{B}^{1/2}} \vee \frac{-n\lambda}{200qm\tau_\infty^2}\right]\end{aligned}$$

Since $n \geqslant 8c^{-2}K^2\mathfrak{a}_{(K)}\tau_\infty^4(1+4\mathfrak{B}) \geqslant 8c^{-2}m^2\mathfrak{a}_m\tau_\infty^4(1+4\mathfrak{B}_q)$, $1 \leqslant m \leqslant K$, letting $\lambda := c\mathfrak{a}_m^{-1/2}$ and using $\mathfrak{a}_m \leqslant \mathfrak{a}_{(K)}$ and $n^{1/2}c \geqslant 4\tau_\infty^2 K\mathfrak{a}_{(K)}^{1/2}$ we obtain

$$\begin{aligned}\mathbf{P}\left(\sup_{t\in\mathcal{B}_{m^2}} |\overline{\nu_t}| \geqslant 2c\mathfrak{a}_m^{-1/2}\right)\\ \leqslant 3\exp\left[\frac{-nc^2}{800\|p_{Z,W}\|_\infty K\mathfrak{a}_{(K)}\tau_\infty^2\mathfrak{B}^{1/2}} \vee \frac{-nc}{200qK\mathfrak{a}_{(K)}^{1/2}\tau_\infty^2}\right]\\ \leqslant 3\exp\left[\frac{-nc^2}{800\|p_{Z,W}\|_\infty K\mathfrak{a}_{(K)}\tau_\infty^2\mathfrak{B}^{1/2}} \vee \frac{-n^{1/2}}{50q}\right]\end{aligned}$$

A combination of the last bound, (D.24) and $\sup_{t\in\mathcal{B}_{m^2}} |\overline{\nu_t}(x)| \geqslant \|[\Xi]_{\underline{m}}\|_s$ implies the assertion, which completes the proof. $\square$

E Proof of Theorem 4.1

Let us first recall notations and gather preliminary results used in the sequel. Keeping in mind the notations given in (2.3) and (2.4) we assume throughout this section $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$ and use in addition to (A.1) for all $m \geqslant 1$ and $n \geqslant 1$

$$\begin{aligned} &\Delta_m^{\mathfrak{t}} = \Delta_m(\mathfrak{t}),\ \Lambda_m^{\mathfrak{t}} = \Lambda_m(\mathfrak{t}),\ \delta_m^{\mathfrak{t}} = m\Delta_m^{\mathfrak{t}}\Lambda_m^{\mathfrak{t}}, \\ &\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}-} = \mathrm{M}_n(4D^2\mathfrak{t}), \quad \mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+} = \mathrm{M}_n(\mathfrak{t}/(4d^2)). \end{aligned} \tag{E.1}$$

Recall that under Assumption A.5 for all $m \geqslant 1$ it holds $D^{-2} \leqslant \mathfrak{t}_m^{-1}\|[T]_{\underline{m}}^{-1}\|_s^2 \leqslant D^2$, $D^{-2} \leqslant \Delta_m^T/\Delta_m^{\mathfrak{t}} \leqslant D^2$, $(1+2\log D)^{-1} \leqslant \Lambda_m^T/\Lambda_m^{\mathfrak{t}} \leqslant (1+2\log D)$, and $D^{-2}(1+2\log D)^{-1} \leqslant \delta_m^T/\delta_m^{\mathfrak{t}} \leqslant D^2(1+2\log D)$ as well as $\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}-} \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} \leqslant \mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+}$, for all $n \geqslant 1$. Furthermore, the elements of $\mathcal{F}_{\mathfrak{f}}^r$ are bounded uniformly, that is, $\|\phi\|_\infty^2 \leqslant \|\sum_{j\geqslant 1}\mathfrak{f}_j u_j^2\|_\infty \|\phi\|_{\mathfrak{f}}^2 \leqslant \tau_{\mathfrak{f},\infty}^2 r^2$ for all $\phi \in \mathcal{F}_{\mathfrak{f}}^r$. The last estimate is used in Lemma B.9 in the Appendix B to show that for all $f \in \mathcal{F}_{\mathfrak{f}}^r$ and $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$ the approximation f_m satisfies $\mathfrak{f}_m^{-1}\mathfrak{b}_m^2(f) \leqslant 4D^4r^2$, $\|f - f_m\|_\infty \leqslant 2D^2\tau_{\mathfrak{f},\infty} r$ and $\|f_m\|_Z^2 \leqslant 4D^4r^2$. Thereby, setting $\Gamma_2^{\mathfrak{f}} := \|\mu\|_{Z,W}^2 \vee 4D^4r^2$ and $\Gamma_2^{\mathfrak{f}} := \|\mu\|_\infty + (1+2D^2)\tau_{\mathfrak{f},\infty} r$ the Assumption A.3 (b) holds with $\Gamma_2^f := \Gamma_2^{\mathfrak{f}}$ and $\Gamma_\infty^f := \Gamma_\infty^{\mathfrak{f}}$ uniformly for all $f \in \mathcal{F}_{\mathfrak{f}}^r$ and $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$. The proof follows along the lines of the proof of Theorem 3.1 given in Appendix C. We shall prove below the Propositions E.1 and E.2 which are used in the proof of Theorem 4.1. In the proof the propositions we refer to the three technical Lemma C.4, C.5 and E.3 which are shown in Appendix C and the end of this section. Moreover, we make use of functions $\Psi^{\mathfrak{t}}, \Phi_{3n}^{\mathfrak{t}}, \Phi_{4n}^{\mathfrak{t}}, \Phi_{5n}^{\mathfrak{t}} : \mathbb{R}_+ \to \mathbb{R}$ defined by

$$\begin{aligned} &\Psi^{\mathfrak{t}}(x) = D^2 \textstyle\sum_{m\geqslant 1} x\mathfrak{t}_m \exp(-m\Lambda_m^{\mathfrak{t}}/(6(1+2\log D)x)), \\ &\Phi_{3n}^{\mathfrak{t}}(x) = n^3 \exp(-n(\Delta_{\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+}}^{\mathfrak{t}})^{-1}/(25600D^2x^2)), \\ &\Phi_{4n}^{\mathfrak{t}}(x) = n^3 \exp(-n(\Delta_{\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+}+1}^{\mathfrak{t}})^{-1}/(6400D^2x)) \\ &\Phi_{5n}^{\mathfrak{t}}(x) = xn\exp(-\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+}\log(n)/(6x)). \end{aligned} \tag{E.2}$$

Note that each function in (E.2) is non decreasing in x and for all $x > 0$, $\Psi^{\mathfrak{t}}(x) < \infty$, $\Psi^T(x) \leqslant \Psi^{\mathfrak{t}}(x)$, $\Phi_{3n}^T(x) \leqslant \Phi_{3n}^{\mathfrak{t}}(x)$ and $\Phi_{4n}^T(x) \leqslant \Phi_{4n}^{\mathfrak{t}}(x)$ with Ψ^T, Φ_{3n}^T and Φ_{4n}^T as in (C.1). Moreover, if $\log(n)(\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+}+1)^2\Delta_{\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+}+1}^T = o(n)$ as $n \to \infty$ then there exists an integer n_o such that

$$1 \geqslant \sup_{n\geqslant n_o}\left\{1024\tau_\infty^4 D^2(1+\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2(\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+}+1)^2\Delta_{\mathrm{M}_{\mathrm{n}}^{\mathfrak{t}+}+1}^{\mathfrak{t}} n^{-1}\right\}, \tag{E.3}$$

and we have also for all $x > 0$, $\Phi_{3n}^{\mathfrak{t}}(x) = o(1)$, $\Phi_{4n}^{\mathfrak{t}}(x) = o(1)$ and $\Phi_{5n}^{\mathfrak{t}}(x) = o(1)$ as $n \to \infty$. Consequently, considering Φ_{1n} and Φ_{2n} as in (C.1) under Assumption A.1 and

A.2 there exists a finite constant $\Sigma^{\mathfrak{f}}$ such that for all $n \geqslant 1$,

$$\Sigma^{\mathfrak{f}} \geqslant \Big\{ n_o^2 \bigvee n^3 \exp(-n^{1/6}/50) \bigvee \Psi^{\mathfrak{t}}\Big(1+\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi\Big) \bigvee \Phi_{1n}\Big(1+\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi\Big) \bigvee \Phi_{2n}\Big(1+\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi\Big) \\ \bigvee \Phi^{\mathfrak{t}}_{5n}\Big(1+\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi\Big) \bigvee \Phi^{\mathfrak{t}}_{3n}(1+\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi) \bigvee \Phi^{\mathfrak{t}}_{4n}(\|p_{Z,W}\|_\infty) \\ \bigvee \mathbb{E}(\xi/\sigma_\xi)^8 \bigvee (\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^8 \bigvee (\tau_\infty/\sigma_\xi)^2 \mathbb{E}(\xi/\sigma_\xi)^{12} \Big\}. \quad \text{(E.4)}$$

PROOF OF THEOREM 4.1. We start the proof considering the elementary identity (C.5) given in the proof of Theorem 3.1 where we bound the two rhs terms separately. The second rhs term we bound with help of Proposition E.2. Thereby, there exists a numerical constant C such that for all $f \in \mathcal{F}^r_{\mathfrak{f}}$ hold

$$\mathbb{E}\|\widehat{f}_{\widehat{m}} - f\|^2_Z \leqslant \mathbb{E}\left(\mathbb{1}_{\mathcal{E}_n}\|\widehat{f}_{\widehat{m}} - f\|^2_Z\right) + C\, n^{-1}\, \tau^2_\infty(1+\sigma^2_\xi + \Gamma^{\mathfrak{f}}_2)\Sigma^{\mathfrak{f}}. \quad \text{(E.5)}$$

Consider the first rhs term. On the event $\mathcal{E}_n$ the upper bound given in (C.4) implies

$$\|\widehat{f}_{\widehat{m}} - f\|^2_Z \mathbb{1}_{\mathcal{E}_n} \leqslant 582 \min_{1\leqslant m\leqslant \mathrm{M}^{\mathrm{T}-}_{\mathrm{n}}}\{[\mathfrak{b}^2_m(f)\vee \mathrm{pen}_m]\} + 42 \max_{1\leqslant k\leqslant \mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}}\Big(\|\widehat{f}_k - f_k\|^2_Z - \mathrm{pen}_k/6\Big)_+.$$

Keeping in mind that $\mathrm{pen}_k = 144\tau^2_\infty\sigma^2_k\delta^T_k n^{-1}$ with $\delta^T_k = k\Lambda^T_k\Delta^T_k$ and $\sigma^2_k \leqslant 2(\sigma^2_\xi + 3\Gamma^{\mathfrak{f}}_2)$ we derive in Proposition E.1 below an upper bound for the expectation of the second rhs term, the remainder term, in the last display. Thereby, from $\mathrm{M}^{\mathrm{T}-}_{\mathrm{n}} \geqslant \mathrm{M}^{\mathrm{t}-}_{\mathrm{n}}$, $\mathfrak{b}^2_m(f) \leqslant \mathfrak{f}_m 4D^4r^2$ and $\delta^T_k \leqslant D^2(1+2\log D)\delta^{\mathfrak{t}}_k$ there exists a numerical constant C such that for all $f \in \mathcal{F}^r_{\mathfrak{f}}$

$$\mathbb{E}\Big(\mathbb{1}_{\mathcal{E}_n}\|\widehat{f}_{\widehat{m}} - f\|^2_Z\Big) \leqslant C\, \tau^2_\infty D^4(r^2 + \sigma^2_\xi + \Gamma^{\mathfrak{f}}_2)\{\min_{1\leqslant m\leqslant \mathrm{M}^{\mathrm{t}-}_{\mathrm{n}}}\{[\mathfrak{f}_m \vee n^{-1}\delta^{\mathfrak{t}}_m]\} + n^{-1}\,\Sigma^{\mathfrak{f}}\}.$$

Replacing in (E.5) the first rhs by the last upper bound we obtain the assertion of the theorem, which completes the proof. □

PROPOSITION E.1. *Under the assumptions of Theorem 4.1 there exists a numerical constant C such that for all $n \geqslant 1$*

$$\mathbb{E}\Big\{\max_{1\leqslant k\leqslant \mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}}\Big(\|\widehat{f}_m - f_m\|^2_Z - 24\tau^2_\infty m n^{-1}\sigma^2_m\Lambda^T_m\Delta^T_m\Big)_+\Big\} \leqslant Cn^{-1}\tau^2_\infty(1+\sigma^2_\xi+\Gamma^{\mathfrak{f}}_2)\Sigma^{\mathfrak{f}}.$$

PROOF OF PROPOSITION E.1. We start the proof with an upper bound similar to (C.7) using $\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}} \leqslant \mathrm{M}^{\mathrm{t}+}_{\mathrm{n}}$, that is,

$$\begin{aligned}
&\mathbb{E}\Big\{\max_{1\leqslant m\leqslant \mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}}\Big(\|\widehat{f}_m - f_m\|^2_Z - 24\tau^2_\infty m n^{-1}\sigma^2_m\Lambda^T_m\Delta^T_m\Big)_+\Big\} \\
&\qquad \leqslant 2\mathbb{E}\Big\{\max_{1\leqslant m\leqslant \mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}}\|[T]^{-1}_{\underline{m}}\|^2_s\Big(\|[V]_{\underline{m}}\|^2 - 12\tau^2_\infty m n^{-1}\sigma^2_m\Lambda^T_m\Big)_+\Big\} \\
&\qquad\quad + \mathbb{E}\Big\{n\Big(\|[V]_{\underline{\mathrm{M}^{\mathrm{t}+}_{\mathrm{n}}}}\|^2 - 12\tau^2_\infty\,\mathrm{M}^{\mathrm{t}+}_{\mathrm{n}}\,n^{-1}\sigma^2_{\mathrm{M}^{\mathrm{t}+}_{\mathrm{n}}}\log(n)\Big)_+\Big\} \\
&\qquad\quad + 12\tau^2_\infty\,\mathrm{M}^{\mathrm{t}+}_{\mathrm{n}}\,\sigma^2_{\mathrm{M}^{\mathrm{t}+}_{\mathrm{n}}}\log(n)P\Big(\bigcup_{k=1}^{\mathrm{M}^{\mathrm{t}+}_{\mathrm{n}}}\mho^c_k\Big) + \max_{1\leqslant m\leqslant \mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}}\|f_m\|^2_Z P\Big(\bigcup_{k=1}^{\mathrm{M}^{\mathrm{T}+}_{\mathrm{n}}}\Omega^c_k\Big) \quad \text{(E.6)}
\end{aligned}$$

where we bound separately each of the four rhs terms. In order to bound the first and second rhs term we employ (C.14) in Lemma C.4 with $K = \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}$ and $K = \mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}$, and sequence $\mathfrak{a} = (\mathfrak{a}_m)_{m\geqslant 1}$ given by $\mathfrak{a}_m = \|[T]_{\underline{m}}^{-1}\|_s^2$ and $\mathfrak{a}_m = n\mathbb{1}_{\{m\,=\,\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}\}}$, respectively. Keeping in mind the definition of $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}$, $\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}$ and $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+} \leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{t}+} \leqslant \lfloor n^{1/4}\rfloor$, and hence in both cases $\mathfrak{a}_{(K)}K^2 \leqslant n^{3/2}$, there exists a numerical constant $C > 0$ such that

$$\mathbb{E}\Big\{\max_{1\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\Big(\|\widehat{f}_m - f_m\|_Z^2 - 24\tau_\infty^2 mn^{-1}\sigma_m^2\Lambda_m^T\Delta_m^T\Big)_+\Big\}$$
$$\leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Psi^T(1+\Gamma_\infty^f/\sigma_\xi)+\sigma_\xi^2\Phi_{2n}(1+\Gamma_\infty^f/\sigma_\xi)+\sigma_\xi^2\Phi_{5n}^{\mathrm{t}}(1+\Gamma_\infty^f/\sigma_\xi)+\mathbb{E}(\xi/\sigma_\xi)^{12}\Big\}$$
$$+\,6\tau_\infty^2\,\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}\,\sigma_{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}^2\log(n)P\Big(\bigcup_{k=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\mho_k^c\Big) + \max_{1\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\|f_m\|_Z^2 P\Big(\bigcup_{k=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\Omega_k^c\Big)$$

with Ψ^T, Φ_{2n} as in (C.1), and Φ_{5n}^{t} as in (E.2). Taking into account that $\Psi^T(x) \leqslant \Psi^{\mathrm{t}}(x)$ and that Assumption A.3 (b) holds with $\Gamma_2^f := \Gamma_2^{\mathfrak{f}}$ and $\Gamma_\infty^f := \Gamma_\infty^{\mathfrak{f}}$ uniformly for all $f \in \mathcal{F}_{\mathfrak{f}}^r$ and $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$, it follows that

$$\mathbb{E}\Big\{\max_{1\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\Big(\|\widehat{f}_m - f_m\|_Z^2 - 24\tau_\infty^2 mn^{-1}\sigma_m^2\Lambda_m^T\Delta_m^T\Big)_+\Big\}$$
$$\leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Psi^{\mathrm{t}}(1+\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)+\sigma_\xi^2\Phi_{2n}(1+\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)+\sigma_\xi^2\Phi_{5n}^{\mathrm{t}}(1+\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)+\mathbb{E}(\xi/\sigma_\xi)^{12}\Big\}$$
$$+\,6\tau_\infty^2\,\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}\,\sigma_{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}^2\log(n)P\Big(\bigcup_{k=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}\mho_k^c\Big) + \max_{1\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\|f_m\|_Z^2 P\Big(\bigcup_{k=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\Omega_k^c\Big)$$

Exploiting that $\sigma_m^2 \leqslant 2(\sigma_\xi^2 + 3\Gamma_2^{\mathfrak{f}})$, $\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}\log(n) \leqslant n$ and $\max_{1\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\|f_m\|_Z^2 \leqslant \Gamma_2^{\mathfrak{f}}$, replacing the probability $P(\bigcup_{k=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}\Omega_k^c)$ and $P(\bigcup_{k=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}\mho_k^c)$ by its upper bound given in (E.11) and (E.8) in Lemma E.3, respectively, and employing the definition of $\Sigma^{\mathfrak{f}}$ as in (E.4) we obtain the result of the proposition, which completes the proof. □

Proposition E.2. *Under the assumptions of Theorem 4.1 there exists a numerical constant C such that for all $n \geqslant 1$*

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2\mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant C\,n^{-1}\,\tau_\infty^2(1+\sigma_\xi^2+\Gamma_2^{\mathfrak{f}})\Sigma^{\mathfrak{f}}.$$

Proof of Proposition E.2. Following line by line the proof of Proposition C.2 for $M := \lfloor n^{1/4}\rfloor$ there exists a numerical constant $C > 0$ such that

$$\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2\mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Phi_{1n}\Big(1+\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi\Big)+\sigma_\xi^2\Phi_{2n}\Big(1+\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi\Big)+\mathbb{E}(\xi/\sigma_\xi)^{12}\Big\}$$
$$+\,\{36\tau_\infty^2 M\sigma_M^2\log(n) + 6\Gamma_2^{\mathfrak{f}}\}P(\mathcal{E}_n^c)$$

with Φ_{1n} and Φ_{2n} as in (C.1). Exploiting further the definition of $\Sigma^{\mathfrak{f}}$ as in (E.4) and that $\sigma_M^2 \leqslant 2\{\sigma_\xi^2 + 3\Gamma_2^{\mathfrak{f}}\}$ and $M\log(n) \leqslant n$ the result of the proposition follows now by replacing the probability $P(\mathcal{E}_n^c)$ by its upper bound given in (E.10) in Lemma E.3, which completes the proof. □

Lemma E.3. *Under the assumptions of Theorem 4.1 there exists a numerical constant C such that for all $n \geqslant 1$*

$$\mathbf{P}\Big(\mathcal{A}_n^c\Big) = \mathbf{P}\Big(\{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}^c\Big) \leqslant C\, \Sigma^{\mathfrak{f}}\, n^{-2}, \tag{E.7}$$

$$\mathbf{P}\Big(\mathcal{B}_n^c\Big) \leqslant \mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_n^{\mathfrak{t}+}+1} \mho_m^c\Big) \leqslant C\, \Sigma^{\mathfrak{f}}\, n^{-2}, \tag{E.8}$$

$$\mathbf{P}\Big(\mathcal{C}_n^c\Big) \leqslant C\, \Sigma^{\mathfrak{f}}\, n^{-2}, \tag{E.9}$$

$$\mathbf{P}\Big(\mathcal{E}_n^c\Big) \leqslant C\, \Sigma^{\mathfrak{f}}\, n^{-2}, \tag{E.10}$$

$$\mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_n^{T+}} \Omega_m\Big) \leqslant C\, \Sigma^{\mathfrak{f}}\, n^{-2}. \tag{E.11}$$

Proof of Lemma E.3. The proof of (E.7) follows line by line the proof of (C.9) in Lemma C.3 using the definition of $\Sigma^{\mathfrak{f}}$ as in (E.4) rather Σ^f than (C.3) and hence we omit the details. Consider (E.8)–(E.10). Let $\mathfrak{a}$ be a sequence given by $\mathfrak{a}_m = \|[T]_{\underline{m}}^{-1}\|_s^2$ where $\mathfrak{a}_{(m)} = \Delta_m^T$ and n_o an integer satisfying (E.3) uniformly for all $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$ and $f \in \mathcal{F}_{\mathfrak{f}}^r$, that is, $n \geqslant 1024\tau_\infty^4 D^2(1+\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2(\mathrm{M}_n^{\mathfrak{t}+}+1)^2\Delta^{\mathfrak{t}}_{\mathrm{M}_n^{\mathfrak{t}+}+1} \geqslant 1024\tau_\infty^4(1+\Gamma_\infty^f/\sigma_\xi)^2(\mathrm{M}_n^{T+}+1)^2\Delta^T_{\mathrm{M}_n^{T+}+1}$ for all $n > n_o$ by construction. We distinguish in the following the cases $n \leqslant n_o$ and $n > n_o$. Consider (E.8). Following line by line the proof of (C.10) together with $\Phi_{4n}^T(x) \leqslant \Phi_{4n}^{\mathfrak{t}}(x)$ and $\mathrm{M}_n^{\mathfrak{t}+}+1 \leqslant n$ we have $\mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_n^{\mathfrak{t}+}+1} \mho_m^c\Big) \leqslant 3\, n^{-2}\, \Phi_{4n}^{\mathfrak{t}}(\|p_{Z,W}\|_\infty) \vee \{n^3\exp(-n^{1/2}/50)\}$. By combination of the two cases and employing the definition of $\Sigma^{\mathfrak{f}}$ given in (E.4) we obtain (E.8). The proof of (E.9) follows line by line the proof of (C.11) in Lemma C.3. Exploiting $\Phi_{3n}^T(x) \leqslant \Phi_{3n}^{\mathfrak{t}}(x)$ we obtain $\mathbf{P}\Big(\mathcal{C}_n^c\Big) \leqslant 3\{n^3\exp(-n^{1/6}/50)\} \vee \Phi_{3n}^{\mathfrak{t}}(1+\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)n^{-2} + 32(\tau_\infty^2/\sigma_\xi^2)\mathbb{E}(\xi/\sigma_\xi)^{12}n^{-2}$ The assertion (E.9) follows employing the definition of $\Sigma^{\mathfrak{f}}$ given in (E.4). Consider (E.10). Due to Lemma B.7 it holds $\mathbf{P}\Big(\mathcal{E}_n^c\Big) \leqslant \mathbf{P}\Big(\mathcal{A}_n^c\Big) + \mathbf{P}\Big(\mathcal{B}_n^c\Big) + \mathbf{P}\Big(\mathcal{C}_n^c\Big)$. Therefore, the assertion (E.10) follows from (E.7)–(E.9). The proof of (E.11) follows in same manner as the proof of (C.13), and we omit the details, which completes the proof. □

F Proof of Theorem 4.3

The proof follows along the lines of the proof of Theorem 3.2 given in Appendix D. We shall prove below the Propositions F.1 and F.2 which are used in the proof of Theorem 4.3. In the proof the propositions we refer to the three technical Lemma D.4, D.5 and F.3 which are shown in Appendix D and the end of this section. Moreover, we make use of functions $\Psi^{\mathfrak{t}}, \Phi^{\mathfrak{t}}_{3n}, \Phi^{\mathfrak{t}}_{4n}, \Phi^{\mathfrak{t}}_{5n} : \mathbb{R}_+ \to \mathbb{R}$ defined by

$$\begin{aligned}
&\Psi^{\mathfrak{t}}(x) = D^2 \textstyle\sum_{m\geqslant 1} x m^{1/2} \mathfrak{t}_m \exp(-m^{1/2}\Lambda^{\mathfrak{t}}_m/(48(1+2\log D)x)),\\
&\Phi^{\mathfrak{t}}_{3n}(x) = n^3 \exp(-n(\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}})^{-1/2}(\Delta^{\mathfrak{t}}_{\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}})^{-1}/(204800D^2x)),\\
&\Phi^{\mathfrak{t}}_{4n}(x) = n^3 \exp(-n(\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1)^{-1}(\Delta^{\mathfrak{t}}_{\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1})^{-1}/(51200D^2x))\\
&\Phi^{\mathfrak{t}}_{5n}(x) = xn \exp(-(\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}})^{1/2}\log(n)/(48x)).
\end{aligned} \tag{F.1}$$

Note that each function in (F.1) is non decreasing in x and for all $x > 0$, $\Psi^{\mathfrak{t}}(x) < \infty$, $\Psi^T(x) \leqslant \Psi^{\mathfrak{t}}(x)$, $\Phi^T_{3n}(x) \leqslant \Phi^{\mathfrak{t}}_{3n}(x)$ and $\Phi^T_{4n}(x) \leqslant \Phi^{\mathfrak{t}}_{4n}(x)$ with Ψ^T, Φ^T_{3n} and Φ^T_{4n} as in (D.1). Moreover, if $\log(n)(\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1)^2\Delta^{\mathfrak{t}}_{\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1} = o(n)$ as $n \to \infty$ then there exists an integer n_o such that

$$1 \geqslant \sup_{n\geqslant n_o} \Big\{1024\tau_\infty^4 D^2(6+8(\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\mathfrak{B})(\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1)^2\Delta^{\mathfrak{t}}_{\mathrm{M}^{\mathfrak{t}+}_{\mathrm{n}}+1}n^{-1}\Big\}, \tag{F.2}$$

and we have also for all $x > 0$, $\Phi^{\mathfrak{t}}_{3n}(x) = o(1)$, $\Phi^{\mathfrak{t}}_{4n}(x) = o(1)$ and $\Phi^{\mathfrak{t}}_{5n}(x) = o(1)$ as $n \to \infty$. Consequently, considering Φ_{1n} and Φ_{2n} as in (D.1) under Assumption A.1 and A.2 there exists a finite constant $\Sigma^{\mathfrak{f}}$ such that for all $n \geqslant 1$,

$$\begin{aligned}
\Sigma^{\mathfrak{f}} \geqslant \Big\{& n_o^2 \bigvee \Psi^{\mathfrak{t}}\Big(1+(\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big) \bigvee \Phi_{1n}\Big(1+(\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big) \bigvee \Phi_{2n}\Big(1+\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi\Big)\\
&\bigvee \Phi^{\mathfrak{t}}_{3n}(1+(\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}) \bigvee \Phi^{\mathfrak{t}}_{4n}(\|p_{Z,W}\|_\infty\mathfrak{B}^{1/2}\tau_\infty^2) \bigvee \Phi^{\mathfrak{t}}_{5n}\Big(1+(\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big)\\
&\bigvee \mathbb{E}(\xi/\sigma_\xi)^8 \bigvee (\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^8\mathfrak{B} \bigvee (\tau_\infty/\sigma_\xi)^2\mathbb{E}(\xi/\sigma_\xi)^{12}\Big\}.
\end{aligned} \tag{F.3}$$

Proof of Theorem 4.3. We start the proof considering the elementary identity (C.5) given in the proof of Theorem 3.1 where we bound the two rhs terms separately. The second rhs term we bound with help of Proposition F.2. Thereby, there exists a numerical constant C such that for all $f \in \mathcal{F}^r_{\mathfrak{f}}$ hold

$$\begin{aligned}
&\mathbb{E}\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \leqslant \mathbb{E}\left(\mathbb{1}_{\mathcal{E}_n}\|\widehat{f}_{\widehat{m}} - f\|_Z^2\right)\\
&+ C\, n^{-1}\, \tau_\infty^2\{\sigma_\xi^2 + \Gamma^{\mathfrak{f}}_2\}(1+(\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\mathfrak{B})[\Sigma^{\mathfrak{f}} \vee n^3\exp(-n^{1/6}q^{-1}/100) \vee n^4q^{-1}\beta_{q+1}].
\end{aligned} \tag{F.4}$$

Consider the first rhs term. On the event $\mathcal{E}_n$ the upper bound given in (C.4) implies

$$\|\widehat{f}_{\widehat{m}} - f\|_Z^2\mathbb{1}_{\mathcal{E}_n} \leqslant 582\{[\mathfrak{b}^2_{m_n^\diamond}(f) \vee \mathrm{pen}_{m_n^\diamond}]\} + 42 \max_{m_n^\diamond\leqslant k\leqslant \mathrm{M}^{T+}_{\mathrm{n}}} \left(\|\widehat{f}_k - f_k\|_Z^2 - \mathrm{pen}_k/6\right)_+.$$

Keeping in mind that $\mathrm{pen}_k = 288\kappa^f_n\tau_\infty^2\sigma_k^2\delta^T_k n^{-1}$ with $\delta^T_k = k\Lambda^T_k\Delta^T_k$, $\kappa^f_n \leqslant 8(1+(\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\mathfrak{B})$ and $\sigma_k^2 \leqslant 2(\sigma_\xi^2 + 3\Gamma^{\mathfrak{f}}_2)$ we derive in Proposition F.1 below an upper bound for the expectation of the second rhs term, the remainder term, in the last display. Thereby, from

$\mathrm{M}_{\mathrm{n}}^{\mathrm{T}-} \geqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{t}-}$, $\mathfrak{b}_m^2(f) \leqslant \mathfrak{f}_m 4D^4r^2$ and $\delta_k^T \leqslant D^2(1+2\log D)\delta_k^{\mathrm{t}}$ there exists a numerical constant C such that for all $f \in \mathcal{F}_{\mathfrak{f}}^r$

$$\mathbb{E}\Big(\mathbb{1}_{\mathcal{E}_n}\|\widehat{f}_{\widehat{m}}-f\|_Z^2\Big) \leqslant C\left\{[\mathfrak{f}_{m_n^\diamond}\vee n^{-1}\delta^{\mathrm{t}}_{m_n^\diamond}]+n^{-1}[\Sigma^{\mathfrak{f}}\vee n^3\exp(-n^{1/6}q^{-1}/100)\vee n^4q^{-1}\beta_{q+1}]\right\}$$
$$\times\, \tau_\infty^2 D^4(r^2+\sigma_\xi^2+\Gamma_2^{\mathfrak{f}})(1+(\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2\mathfrak{B}).$$

Replacing in (F.4) the first rhs by the last upper bound we obtain the assertion of the theorem, which completes the proof. □

Proposition F.1. *Under the assumptions of Theorem 4.3 there exists a numerical constant C such that for all $n \geqslant 1$*

$$\mathbb{E}\Big\{\max_{m_n^\diamond\leqslant k\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\Big(\|\widehat{f}_m-f_m\|_Z^2-48\tau_\infty^2\sigma_m^2\kappa_n^f m\Lambda_m^T\Delta_m^T n^{-1}\Big)_+\Big\}$$
$$\leqslant C\,n^{-1}\,\tau_\infty^2\{\sigma_\xi^2+\Gamma_2^{\mathfrak{f}}\}(1+(\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2\mathfrak{B})[\Sigma^{\mathfrak{f}}\vee n^3\exp(-n^{1/6}q^{-1}/100)\vee n^4q^{-1}\beta_{q+1}].$$

Proof of Proposition F.1. We start the proof with an upper bound similar to (C.7) using $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}$, that is,

$$\begin{aligned}
\mathbb{E}\Big\{\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\Big(\|\widehat{f}_m-f_m\|_Z^2-48\tau_\infty^2\kappa_n^f\sigma_m^2 m\Lambda_m^T\Delta_m^T n^{-1}\Big)_+\Big\}&\\
\leqslant 2\mathbb{E}\Big\{\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\|[T]_{\underline{m}}^{-1}\|_s^2\Big(\|[V]_{\underline{m}}\|^2-24\tau_\infty^2\kappa_n^f\sigma_m^2 m\Lambda_m^T n^{-1}\Big)_+\Big\}&\\
+\mathbb{E}\Big\{n\Big(\|[V]_{\underline{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}}\|^2-24\tau_\infty^2\sigma^2_{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}\kappa_n^f\,\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}\log(n)n^{-1}\Big)_+\Big\}&\\
+24\tau_\infty^2\sigma^2_{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}\kappa_n^f\,\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}\log(n)P\Big(\bigcup_{k=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}\mho_k^c\Big)+\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\|f_m\|_Z^2P\Big(\bigcup_{k=1}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\Omega_k^c\Big)& \quad\text{(F.5)}
\end{aligned}$$

where we bound separately each of the four rhs terms. In order to bound (i) the first and (ii))second rhs term we employ (D.12) in Lemma D.4 with $k=m_n^\diamond$, (i) $K=\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}$ and (ii) $K=\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}$, and sequence $\mathfrak{a}=(\mathfrak{a}_m)_{m\geqslant1}$ given by (i) $\mathfrak{a}_m=\|[T]_{\underline{m}}^{-1}\|_s^2$ and (ii) $\mathfrak{a}_m=n\mathbb{1}_{\{m=\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}\}}$. Keeping in mind the definition of $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}$, $\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}$ and $\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}\leqslant\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}\leqslant\lfloor n^{1/4}\rfloor$, and hence in both cases $\mathfrak{a}_{(K)}K^2\leqslant n^{3/2}$, there exists a numerical constant $C>0$ such that

$$\begin{aligned}
&\mathbb{E}\Big\{\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\Big(\|\widehat{f}_m-f_m\|_Z^2-48\tau_\infty^2\kappa_n^f\sigma_m^2 m\Lambda_m^T\Delta_m^T n^{-1}\Big)_+\Big\}\\
&\leqslant Cn^{-1}\tau_\infty^2\Big\{\sigma_\xi^2\Psi^T\Big(1+(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big)+\sigma_\xi^2\Phi_{2n}\Big(1+\Gamma_\infty^f/\sigma_\xi\Big)\\
&\qquad+\sigma_\xi^2\Phi_{5n}^{\mathrm{t}}\Big(1+(\Gamma_\infty^f/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big)+\sigma_\xi^2(1+\Gamma_\infty^f/\sigma_\xi)^2n^{7/3}q^{-1}\beta_{q+1}+\mathbb{E}(\xi/\sigma_\xi)^6\Big\}\\
&\qquad+24\tau_\infty^2\kappa_n^f\sigma^2_{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}\,\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}\log(n)P\Big(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{t}+}}\mho_k^c\Big)+\max_{m_n^\diamond\leqslant m\leqslant \mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\|f_m\|_Z^2P\Big(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_{\mathrm{n}}^{\mathrm{T}+}}\Omega_k^c\Big)
\end{aligned}$$

with Ψ^T, Φ_{2n} as in (D.1) and $\Phi^{\mathfrak{t}}_{5n}$ as in (F.1). Taking into account that $\Psi^T(x) \leqslant \Psi^{\mathfrak{t}}(x)$ and that Assumption A.3 (b) holds with $\Gamma^f_2 := \Gamma^{\mathfrak{f}}_2$ and $\Gamma^f_\infty := \Gamma^{\mathfrak{f}}_\infty$ uniformly for all $f \in \mathcal{F}^r_{\mathfrak{f}}$ and $T \in \mathcal{T}^{d,D}_{\mathfrak{t}}$, it follows that

$$\begin{aligned}
&\mathbb{E}\Big\{\max_{m_n^\diamond \leqslant m \leqslant \mathrm{M}_n^{\mathrm{T}+}} \Big(\|\widehat{f}_m - f_m\|_Z^2 - 48\tau_\infty^2 \kappa_n^f \sigma_m^2 m \Lambda_m^T \Delta_m^T n^{-1}\Big)_+\Big\} \\
&\leqslant C n^{-1}\tau_\infty^2\Big\{\sigma_\xi^2 \Psi^{\mathfrak{t}}\Big(1 + (\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2 \tau_\infty \mathfrak{B}^{1/2}\Big) + \sigma_\xi^2 \Phi_{2n}\Big(1 + \Gamma^{\mathfrak{f}}_\infty/\sigma_\xi\Big) \\
&\qquad + \sigma_\xi^2 \Phi^{\mathfrak{t}}_{5n}\Big(1 + (\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2 \tau_\infty \mathfrak{B}^{1/2}\Big) + \sigma_\xi^2 (1 + \Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2 n^{7/3} q^{-1} \beta_{q+1} + \mathbb{E}(\xi/\sigma_\xi)^6\Big\} \\
&\qquad + 24\tau_\infty^2 \kappa_n^f \sigma^2_{\mathrm{M}_n^{\mathfrak{t}+}} \mathrm{M}_n^{\mathfrak{t}+} \log(n) P\Big(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_n^{\mathfrak{t}+}} \mho_k^c\Big) + \max_{m_n^\diamond \leqslant m \leqslant \mathrm{M}_n^{\mathrm{T}+}} \|f_m\|_Z^2 P\Big(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_n^{\mathrm{T}+}} \Omega_k^c\Big)
\end{aligned}$$

Exploiting that $\sigma_m^2[\mathrm{M}_n^{\mathrm{T}+}] \leqslant 2(\sigma_\xi^2 + 3\Gamma^{\mathfrak{f}}_2)$, $\kappa_n^f \leqslant 8(1 + (\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\mathfrak{B})$, $\mathrm{M}_n^{\mathfrak{t}+}\log(n) \leqslant n$ and $\max_{m_n^\diamond \leqslant m \leqslant \mathrm{M}_n^{\mathrm{T}+}} \|f_m\|_Z^2 \leqslant \Gamma^{\mathfrak{f}}_2$, replacing the probability $P(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_n^{\mathrm{T}+}} \Omega_k^c)$ and $P(\bigcup_{k=m_n^\diamond}^{\mathrm{M}_n^{\mathfrak{t}+}} \mho_k^c)$ by its upper bound given in (F.10) and (F.7) in Lemma F.3, respectively, and employing the definition of $\Sigma^{\mathfrak{f}}$ as in (F.3) we obtain the result of the proposition, which completes the proof. □

Proposition F.2. *Under the assumptions of Theorem 4.3 there exists a numerical constant C such that for all $1 \leqslant q \leqslant n$*

$$\begin{aligned}
\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant C\, n^{-1}[\Sigma^{\mathfrak{f}} \vee n^3 \exp(-n^{1/6} q^{-1}/100) \vee n^4 q^{-1} \beta_{q+1}] \\
\times\ \tau_\infty^2\{\sigma_\xi^2 + \Gamma^{\mathfrak{f}}_2\}(1 + (\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\mathfrak{B}).
\end{aligned}$$

Proof of Proposition F.2. Following line by line the proof of Proposition D.2 with $M := \lfloor n^{1/4} \rfloor$, $\Gamma^f_\infty := \Gamma^{\mathfrak{f}}_\infty$ and $\kappa_n^f := 8(1 + (\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\mathfrak{B})$ we have

$$\begin{aligned}
\mathbb{E}\Big(\|\widehat{f}_{\widehat{m}} - f\|_Z^2 \mathbb{1}_{\mathcal{E}_n^c}\Big) \leqslant C n^{-1}\tau_\infty^2\Big\{\sigma_\xi^2 \Phi_{1n}\Big(1 + (\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2}\Big) + \sigma_\xi^2\Phi_{2n}\Big(1 + \Gamma^{\mathfrak{f}}_\infty/\sigma_\xi\Big) \\
+ \sigma_\xi^2(1 + \Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2 n^{7/3} q^{-1}\beta_{q+1} + \mathbb{E}(\xi/\sigma_\xi)^{12}\Big\} \\
+ \{72\tau_\infty^2 M \kappa_n^f \sigma_M^2 \log(n) + 6\Gamma^{\mathfrak{f}}_2\} P(\mathcal{E}_n^c)
\end{aligned}$$

with Φ_{1n} and Φ_{2n} as in (D.1). Exploiting further the definition of $\Sigma^{\mathfrak{f}}$ as in (F.3) and that $\sigma_M^2 \leqslant 2\{\sigma_\xi^2 + 3\Gamma^{\mathfrak{f}}_2\}$, $\kappa_n^f = 8(1 + (\Gamma^{\mathfrak{f}}_\infty/\sigma_\xi)^2\mathfrak{B})$ and $M\log(n) \leqslant n$ the result of the proposition follows now by replacing the probability $P(\mathcal{E}_n^c)$ by its upper bound given in (F.9) in Lemma F.3, which completes the proof. □

Lemma F.3. *Under the assumptions of Theorem 4.3 there exists a numerical constant*

C *such that for all* $1 \leqslant q \leqslant n$

$$\mathbf{P}\Big(\mathcal{A}_n^c\Big) = \mathbf{P}\Big(\{1/2 \leqslant \widehat{\sigma}_Y^2/\sigma_Y^2 \leqslant 3/2\}^c\Big) \leqslant C\ \Sigma^{\mathfrak{f}}\ n^{-2}, \tag{F.6}$$

$$\mathbf{P}\Big(\mathcal{B}_n^c\Big) \leqslant \mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_n^{\mathfrak{t}+}+1} \mho_m^c\Big) \leqslant C\ [\Sigma^{\mathfrak{f}} \vee n^3 \exp(-n^{1/2}q^{-1}/50) \vee n^4 q^{-1}\beta_{q+1}]\ n^{-2}, \tag{F.7}$$

$$\mathbf{P}\Big(\mathcal{C}_n^c\Big) \leqslant C\ [\Sigma^{\mathfrak{f}} \vee n^3 \exp(-n^{1/6}q^{-1}/100) \vee n^4 q^{-1}\beta_{q+1}]\ n^{-2}, \tag{F.8}$$

$$\mathbf{P}\Big(\mathcal{E}_n^c\Big) \leqslant C\ [\Sigma^{\mathfrak{f}} \vee n^3 \exp(-n^{1/6}q^{-1}/100) \vee n^4 q^{-1}\beta_{q+1}]\ n^{-2}, \tag{F.9}$$

$$\mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_n^{\mathrm{T}+}} \Omega_m\Big) \leqslant C\ [\Sigma^{\mathfrak{f}} \vee n^3 \exp(-n^{1/2}q^{-1}/50) \vee n^4 q^{-1}\beta_{q+1}]\ n^{-2}. \tag{F.10}$$

Proof of Lemma F.3. The proof of (F.6) follows line by line the proof of (D.6) in Lemma D.3 using the definition of $\Sigma^{\mathfrak{f}}$ as in (F.3) rather Σ^f than (D.3) and hence we omit the details. Consider (F.7)–(F.9). Let $\mathfrak{a}$ be a sequence given by $\mathfrak{a}_m = \|[T]_{\underline{m}}^{-1}\|_s^2$ where $\mathfrak{a}_{(m)} = \Delta_m^T$ and n_o an integer satisfying D.2 uniformly for all $T \in \mathcal{T}_{\mathfrak{t}}^{d,D}$ and $f \in \mathcal{F}_{\mathfrak{f}}^r$, that is, $n \geqslant 1024\tau_\infty^4 D^2(6 + 8(\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2\mathfrak{B})(\mathrm{M}_n^{\mathfrak{t}+}+1)^2\Delta^{\mathfrak{t}}_{\mathrm{M}_n^{\mathfrak{t}+}+1} \geqslant 1024\tau_\infty^4(6 + 8(\Gamma_\infty^f/\sigma_\xi)^2\mathfrak{B})(\mathrm{M}_n^{\mathrm{T}+}+1)^2\Delta^T_{\mathrm{M}_n^{\mathrm{T}+}+1}$ for all $n > n_o$ by construction. We distinguish in the following the cases $n \leqslant n_o$ and $n > n_o$. Consider (F.7). Following line by line the proof of (D.7) together with $\Phi_{4n}^T(x) \leqslant \Phi_{4n}^{\mathfrak{t}}(x)$ and $\mathrm{M}_n^{\mathfrak{t}+}+1 \leqslant n$ we have $\mathbf{P}\Big(\bigcup_{m=1}^{\mathrm{M}_n^{\mathfrak{t}+}+1}\mho_m^c\Big) \leqslant 6\ n^{-2}\ \Big\{\Phi_{4n}^{\mathfrak{t}}(\tau_\infty^2\|p_{Z,W}\|_\infty\mathfrak{B}^{1/2}) \vee n^3\exp(-n^{1/2}q^{-1}/50) \vee n^4 q^{-1}\beta_{q+1}\Big\}$. By combination of the two cases and employing the definition of $\Sigma^{\mathfrak{f}}$ given in (F.3) we obtain (F.7). The proof of (F.8) follows line by line the proof of (D.8) in Lemma D.3. Exploiting $\Phi_{3n}^T(x) \leqslant \Phi_{3n}^{\mathfrak{t}}(x)$ we obtain $\mathbf{P}\Big(\mathcal{C}_n^c\Big) \leqslant 6\{n^3\exp(-n^{1/6}q^{-1}/100)\} \vee \Phi_{3n}^{\mathfrak{t}}(1+(\Gamma_\infty^{\mathfrak{f}}/\sigma_\xi)^2\tau_\infty\mathfrak{B}^{1/2})n^{-2} + nq^{-1}\beta_{q+1} + 64(\tau_\infty^2/\sigma_\xi^2)\mathbb{E}(\xi/\sigma_\xi)^{12}n^{-2}$. The assertion (F.8) follows employing the definition of $\Sigma^{\mathfrak{f}}$ given in (F.3). Consider (F.9). Due to Lemma B.7 it holds $\mathbf{P}\Big(\mathcal{E}_n^c\Big) \leqslant \mathbf{P}\Big(\mathcal{A}_n^c\Big) + \mathbf{P}\Big(\mathcal{B}_n^c\Big) + \mathbf{P}\Big(\mathcal{C}_n^c\Big)$. Therefore, the assertion (F.9) follows from (F.6)–(F.8). The proof of (F.10) follows in same manner as the proof of (D.10), and we omit the details, which completes the proof. □

References

C. Ai and X. Chen. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843, 2003.

T. Amemiya. The nonlinear two-stage least-squares estimator. *Journal of Econometrics*, 2(2):105–110, 1974.

N. Asin and J. Johannes. Adaptive nonparametric estimation in the presence of dependence. Technical report, Université catholique de Louvain, 2016. URL http://arxiv.org/abs/1602.00531.

A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields*, 113(3):301–413, 1999.

L. Birgé and P. Massart. From model selection to adaptive estimation. Pollard, David (ed.) et al., Festschrift for Lucien Le Cam: research papers in probability and statistics. New York, NY: Springer. 55-87, 1997.

L. Birgé, P. Massart, et al. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998.

R. Blundell and J. L. Horowitz. A non-parametric test of exogeneity. *The Review of Economic Studies*, 74(4):1035–1058, 2007.

R. Blundell, X. Chen, and D. Kristensen. Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669, 2007.

D. Bosq. *Nonparametric statistics for stochastic processes*. Number 110 in Lecture Notes in Statistics. Springer, New York, 1998.

C. Breunig and J. Johannes. Adaptive estimation of functionals in nonparametric instrumental regression. *Econometric Theory*, pages 1–43, 2015.

C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.*, 52(1):111–128, 2007a.

C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. II. *Teor. Veroyatn. Primen.*, 52(2):336–349, 2007b.

M. Carrasco, J.-P. Florens, and E. Renault. Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. In *Handbook of Econometrics*, volume 6. North Holland, 2006.

M. Carrasco, J.-P. Florens, and E. Renault. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of econometrics*, 6:5633–5751, 2007.

X. Chen and M. Reiß. On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27:497–521, 2011.

F. Comte and J. Johannes. Adaptive functional linear regression. *The Annals of Statistics*, 40(6):2765–2797, 2012.

F. Comte and F. Merlevede. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM: Probability and Statistics*, 6:211–238, 2002.

S. Darolles, Y. Fan, J.-P. Florens, and E. Renault. Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565, 2011.

F. Dunker, J.-P. Florens, T. Hohage, J. Johannes, and E. Mammen. Iterative estimation of solutions to noisy nonlinear operator equations in nonparametric instrumental regression. *Journal of Econometrics*, 178:444–455, 2014.

S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems.* Kluwer Academic, Dordrecht, 2000.

F. Fève and J.-P. Florens. Iterative algorithm for non parametric estimation of the instrumental variables quantiles. *Economics Letters*, 123(3):300–304, 2014.

J.-P. Florens. Inverse problems and structural econometrics. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 2, pages 46–85. Cambridge University Press, 2003.

J.-P. Florens and A. Simoni. Nonparametric estimation of an instrumental regression: a quasi-bayesian approach based on regularized posterior. *Journal of Econometrics*, 170(2):458–475, 2012.

J.-P. Florens, J. Johannes, and S. Van Bellegem. Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, 27(03): 472–496, 2011.

J.-P. Florens, J. Johannes, and S. Van Bellegem. Instrumental regression in partially linear models. *The Econometrics Journal*, 15(2):304–324, 2012.

P. Gagliardini and O. Scaillet. Tikhonov regularization for nonparametric instrumental variable estimators. *Journal of Econometrics*, 167(1):61–75, 2012.

A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39: 1608–1632, 2011.

P. Hall and J. L. Horowitz. Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904–2929, 2005.

M. Hoffmann and M. Reiß. Nonlinear estimation for linear inverse problems with error in the operator. *The Annals of Statistics*, 36(1):310–336, 2008.

J. L. Horowitz. Adaptive nonparametric instrumental variables estimation: Empirical choice of the regularization parameter. *Journal of Econometrics*, 180(2):158–173, 2014.

J. L. Horowitz and S. Lee. Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, 75(4):1191–1208, 2007.

J. Johannes and M. Schwarz. Partially adaptive nonparametric instrumental regression. *Journal of the Indian Statistical Association*, 49(2):149–175, 2011.

J. Johannes, S. Van Bellegem, and A. Vanhems. Iterative regularisation in nonparametric instrumental regression. *Journal of Statistical Planning and Inference*, 143(1):24–39, 2013.

T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060–1077, 2005.

A. Kolmogorov and Y. Rozanov. On the strong mixing conditions for stationary gaussian sequences. *Theory of Probability and its Applications*, 5:204–207, 1960.

O. Lepski. On a problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.*, 35(3):459–470, 1990.

J.-M. Loubes and C. Marteau. Oracle inequalities for instrumental variable regression. Technical report, Toulouse, 2009. URL http://arxiv.org/abs/0901.4321v1.

P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Analysis*, 18:29–37, 1984.

W. K. Newey. Efficient instrumental variables estimation of nonlinear models. *Econometrica: Journal of the Econometric Society*, pages 809–837, 1990.

W. K. Newey and J. L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

A. Pagan and A. Ullah. *Nonparametric econometrics.* Cambridge university press, 1999.

V. V. Petrov. *Limit theorems of probability theory. Sequences of independent random variables.* Oxford Studies in Probability. Clarendon Press., Oxford, 4. edition, 1995.

M. Talagrand. New concentration inequalities in product spaces. *Inventiones mathematicae*, 126:505–563, 1996.

G. Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probability theory and related fields*, 107(4):467–492, 1997.